BEST AVAILABLE COPY

C5170-1 7/15/82

Medical Care

July 1982
Vol. 20 No. 7
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project

Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems

Effect of a WIC Program on Children's Clinic Activity in a Local Health Department

Closure Among U.S. Community Hospitals, 1976–1980: A Description and a Predictive Model

Self-Care Education: Impact on HMO Costs

Reimbursing for Residency Training: How Many Times?

Procedures and Discrepancies of Blood Pressure Measurements in Two Community Health Centers

Effectiveness of an Ambulatory Care Telephone Service in Reducing Drop-In Visits and Improving Satisfaction With Care

PUBLISHED BY
J. B. Lippincott Company · Philadelphia · Toronto

C5170-1

EDITORIAL
BOARD

**Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington**

S. E. BERKI, M.A.
University of Michigan

ROBERT H. BROOK, M.D., SC.D.
University of California
Los Angeles

GORDON H. DE FRIESE, PH.D.
University of North Carolina

EUGENE FEINGOLD, PH.D.
University of Michigan

BRANT E. FRIES, PH.D.
Yale University

THOMAS R. GODKINS, B.S., P.A.
University of Oklahoma

HAROLD HUNTER, DR. P.H.
Washington, D.C.

SHERMAN A. JAMES, PH.D.
University of North Carolina

ANTHONY L. KOMAROFF, M.D.
Brigham and Women's
Hospital, Boston

JAMES P. LOGERFO, M.D., M.P.H.
University of Washington

NANCY MILIO, PH.D.
University of North Carolina

DUNCAN NEUHAUSER, PH.D.
Case Western Reserve University

JOHN E. WARE, JR., PH.D.
The Rand Corporation

ALONZO S. YERBY, M.D.
Harvard University

ALFRED YANKAUER, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. **Address for subscription information, orders or changes of address:** (except Japan) 2350 Virginia Avenue, Hagerstown, MD 21740. In Japan, contact Woodbell Scope, 11-11, Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. **Annual subscription rates:** U.S. $45.00 individual, $50.00 institution; all other countries except Japan, $55.00 individual, $60.00 institution; special student rate, $27.00. Single copies $6.00. Copies will be replaced without charge if the publisher receives a request within 60 days of the mailing date in the U.S. or within 3 months in all other countries. Printed in U.S.A. © Copyright 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, MD, and at additional mailing offices.

POSTMASTER: Send address changes to Medical Care, 2350 Virginia Avenue, Hagerstown, MD 21740.

July 1982
Volume 20, Number 7

CONTENTS

POLICY FORUM

On "A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project." *David Banta* 661

ORIGINAL ARTICLES

A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project. *Thomas E. Kottke, Pekka Puska, Roger Feldman, Jukka T. Salonen and Jaakko Tuomilehto* 663

Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems. *Paul A. Nutting, Barton R. Burkhalter, Duane L. Dietrick and Edward F. Helmick* 676

Effect of a WIC Program on Children's Clinic Activity in a Local Health Department. *Jonathan B. Kotch and David Whiteman* 691

Closure Among U.S. Community Hospitals, 1976–1980: A Description and a Predictive Model. *Ross M. Mullner, Calvin S. Byre, Paul S. Levy and Joseph D. Kubal* 699

Self-Care Education: Impact on HMO Costs. *Donald W. Kemper* 710

Reimbursing for Residency Training: How Many Times? *Mohan L. Garg, Mounir Elkhatib, Warren M. Kleinberg and Jack L. Mulligan* 719

Procedures and Discrepancies of Blood Pressure Measurements in Two Community Health Centers. *Larry W. Scherwitz, Lynn A. Evans, Deborah J. Hennrikus and Carlos Vallbona* 727

Effectiveness of an Ambulatory Care Telephone Service in Reducing Drop-In Visits and Improving Satisfaction With Care. *Clarence F. Stirewalt, Margaret W. Linn, Gustavo Godoy, Felipe Knopka and Bernard S. Linn* 739

BOOK REVIEWS

Antitrust and the Health Care Provider, by Martin J. Thompson. *Michael R. Pollard* 749

Staffing Primary Care in 1990: Physician Replacement and Cost Savings, edited by Jane Cassels Record. *James E. C. Walker* 750

Politics and the Expanding Physician Supply, by Michael L. Millman. *Eugene Feingold* 751

Managing Change and Collaboration in the Health System: The Paradigm Approach, by Alan Sheldon. *Howard S. Zuckerman* 753

MEDICAL CARE
July 1982, Vol. XX, No. 7

Original Articles

A Decline in Earning Losses Associated With a Community-Based Cardiovascular Disease Prevention Project

THOMAS E. KOTTKE, M.D.,[*] PEKKA PUSKA, M.D.,[†‡]
ROGER FELDMAN, PH.D.,[§] JUKKA T. SALONEN, M.D.,[‡]
AND JAAKKO TUOMILEHTO, M.D.[‡]

Permanent and temporary disability pension award data and survey data were used to estimate the lower bound of earning losses related to cardiovascular disease during a community-based cardiovascular disease prevention program in Eastern Finland. Earning losses due to death totaled $39.94 million but were not affected by the project. Earning losses due to permanent disability totalled $29.01 million and were $4.25 million less than expected ($p < 0.025$). Earning losses due to temporary disability were $10.91 million and were not affected by the project. Total earning losses attributable to cardiovascular disease during the project period were $79.86 million, and total decline in earning losses was $4.25 million. Project implementation costs were less than one per cent of total earning losses and were equal to approximately 17 per cent of the decline in these losses attributable to the project. The findings suggest that community-based heart disease prevention programs have the potential of more than paying for themselves through an associated decline in lost earnings.

CARDIOVASCULAR DISEASE (CVD) places a burden on the community by interfering with a person's ability to work. Rice[1] estimated that CVD cost the United States $6.4 billion in 1963, and Cooper and Rice[2] estimated that CVD cost the United States over $40 billion in 1972. They concluded that indirect costs—lost work or livelihood—accounted for 64 per cent of the total costs in 1967 and 73 per cent of the total costs in 1972. Acton[3] estimated that deaths due to heart attacks alone cost the United States $12 billion in lost earnings in 1967. In Finland in 1972, CVD cost $26 million in disability payments (approximately 25 per cent of all disability payments) and an estimated $100 million in lost earnings.[4] In 1973, CVD accounted for almost 15 per cent of all sickness benefit days in England and Wales.[5] In the same year, the proportion of these diseases at-

---

[*] Laboratory of Physiological Hygiene and [§]Center for Health Services Research, School of Public Health, University of Minnesota, Minneapolis, Minnesota.

[†] Epidemiological Research Unit, National Public Health Laboratory of Finland, Helsinki, Finland.

[‡] Coordinating Center of the North Karelia Project, Research Institute of Public Health, University of Kuopio, Kuopio, Finland.

Supported in part by NIH training grant 5T32 HL-07036 and by the Coordinating Center of the North Karelia Project.

Address for reprints: Thomas E. Kottke, M.D., Laboratory of Physiological Hygiene, Stadium Gate 27, 611 Beacon Street SE, Minneapolis, MN 55455.

0025-7079/82/0700/0663/$01.15 © J. B. Lippincott Co.

tributable to modifiable risk factors cost Canada more than $500 million; almost 90 per cent of the cost was due to lost income.[6] However, Weinstein and Stason,[7] in their theoretic cost-benefit analysis, suggested that CVD prevention programs based on hypertension treatment cannot recover their own costs, mostly because they projected the beneficial effects to occur far in the future, while treatment costs begin to accrue immediately.

An intervention program may reduce work loss by decreasing disease prevalence or incidence. Conversely, increased case-finding or increased awareness and apprehension about heart disease might increase work loss and the demand for disability pensions and related social benefits. For these reasons, and because cost-effectiveness has become an issue in public health programs, as part of the North Karelia Project evaluation two questions about circulatory diseases and lost livelihood were addressed: 1) How much did heart and circulatory diseases cost North Karelia in lost earnings during the years of the North Karelia Project? and 2) Were there cost changes associated with the Project?

Project Background

The North Karelia Project was started in Finland in 1972 to carry out a comprehensive community program for CVD control. The target group was a county of 180,000 inhabitants with exceptionally high heart disease rates.[8] A quasi-experimental evaluation design with a matched reference area was used to assess the feasibility and effects of this community-based intervention for the five-year period between

TABLE 1. Selected Social and Economic Indicators in North Karelia, the Reference Area and All Finland During the Project Period

Indicator	Year	North Karelia	Reference Area	All Finland
Population (1,000s)	1972	181	253	4639
	1976	177	251	4726
Per cent men	1972	50.0	49.0	48.3
	1976	49.7	48.9	48.3
Proportion of population of working age (aged 15–64)	1972	66.1	65.8	66.9
	1976	67.5	67.0	67.4
Net migration of population aged 15–44 (%)	1972	−2.9	−1.8	+0.1
	1976	−1.1	−0.6	−0.4
Persons per household	1970	3.4	3.4	3.0
	1975	3.0	3.0	2.7
Persons per room	1970	1.3	1.2	1.0
	1975	1.0	1.0	0.9
Per cent of labor force engaged in farming or forestry	1970	39.4	34.3	20.3
	1976	29.0	25.4	15.0
Taxable annual income (Finnmarks/person)	1972	3824	4156	5771
	1976	8999	9640	12631
Per cent of working age population receiving disability pension	1972	11.4	9.6	6.6
	1976	13.0	11.9	8.1
Per cent of population receiving social assistance	1972	6.3	6.2	5.5
	1976	4.9	4.4	4.1

spring 1972 and spring 1977. The reference area was selected because it was the county in Finland that most resembled North Karelia in age structure, risk factor levels, disease incidence rates and socioeconomic characteristics (Table 1). Before the beginning of the intervention and after the end of the pilot period, separate 6.5 per cent samples of the middle-aged populations (born between the years of 1913 and 1947) of North Karelia and the reference area were asked to answer questions regarding beliefs about CVD, risk-factor-related behaviors and socioeconomic status. Participants also had physical examinations. About 91 per cent of the selected sample participated in the surveys (Table 2).

The project intervention goal was to influence disease rates by reducing the population levels of the major cornary heart disease (CHD) risk factors, i.e., smoking, serum cholesterol and blood pressure. Community-based antismoking health education was provided, general dietary changes were promoted to influence serum cholesterol levels, and systematic hypertension detection, treatment and follow-up were organized. The major intervention elements included public health education, organizing individual and group services, promoting environmental changes, training local personnel, and monitoring the changes and program impact.[9] For efficiency and effectiveness, the activities were integrated into the existing community service structure and social organization. The project team helped to plan the activities, supervise their implementation, train other personnel and

TABLE 2. Response Rates to North Karelia Project Population Surveys

	North Karelia		Reference Area	
	Men	Women	Men	Women
1972 Survey				
Original sample (n)	2576	2539	3709	3639
Per cent excluded*	6.6	4.6	5.7	3.5
Purified sample (n)	2407	2423	3499	3510
Nonparticipation (n)				
Temporarily away	65	33	84	69
Unable to answer†	17	14	26	18
Refused	8	3	2	4
Unknown reason	89	66	228	173
Total	179	116	340	264
Participants (n)	2228	2307	3159	3246
Per cent of purified sample	92.6	95.2	90.3	92.5
1977 Survey				
Original sample (n)	2355	2373	3310	3469
Per cent excluded*	1.6	1.3	1.4	0.9
Purified sample (n)	2317	2342	3263	3439
Nonparticipation (n)				
Temporarily away	35	23	45	51
Unable to answer†	30	14	39	33
Refused	23	19	15	29
Unknown reason	227	165	246	168
Total	315	221	345	281
Participants (n)	2002	2121	2918	3158
Per cent of purified sample	85.4	90.6	89.4	91.8

* For reasons of death or emigration.

† For reasons of illness or mental incapacity.

prepare educational materials. The team also catalyzed program development in the community.

A thorough evaluation has demonstrated the success of this program.[9] The program was associated with risk-factor-level declines of 17 per cent among men and 12 per cent among women after adjustment for simultaneous changes in the reference community.[10]

Disease Prevalence and Income and Labor Force Participation Estimates

In order to estimate the effect of CVD on earning losses, the number of workers with CVD, the average income per worker, unemployment rates and labor force participation rates were first estimated. The methods for estimating these rates are presented in this section.

Disease Prevalence Estimation. In this study, heart disease was considered to include stroke, hypertension, congestive heart failure, myocardial infarction, arrhythmia and angina pectoris. The net change in the prevalence of subjects reporting heart disease diagnosed or treated during the previous year was computed by subtracting the change in reporting in the reference area from the change in reporting in North Karelia:

$$\text{net change in CVD prevalence} = (NK_2 - NK_1) - (R_2 - R_1),$$

where NK refers to North Karelia; R refers to the reference area; 1 refers to the baseline survey; and 2 refers to the terminal survey.

The change in disease prevalence was tested for statistical significance with the normal distribution:

$$Z = \frac{\text{net change in CVD prevalence}}{\sqrt{p_{NK}(1 - p_{NK})(1/n_{NK_1} + 1/n_{NK_2}) + p_R(1 - p_R)(1/n_{R_1} - 1/n_{R_2})}}$$

Income Estimation. Mean expected income for individuals free of disease in 1972 and 1977 was calculated from the survey responses to the question "How big was the total income of your household last year, not excluding taxes?" by subjects who defined themselves as main family financial supporters and who denied having any disease treated or diagnosed by a physician within the previous year. Subjects could make one of nine precoded responses about their income; each response, except those for the highest and lowest income, was recoded to be equivalent to the median income in the response range. The highest income response ("over 32,000 Fmk/year") was recoded as "34,000 Fmk/year." The lowest income response ("below 4,000 Fmk/year") was recoded as "2,000 Fmk/year." Annual income estimates were divided by 250 for estimate of daily income.

Employment and Labor Force Participation Rate Estimation. Unemployment would prevent some of those out of work because of disability from woking even if they were not disabled. Mushkin[11] set a now accepted precedent of assuming full employment (e.g., 96 or 97 per cent) as the proportion of disabled who would be working if not disabled. However, Klarman[12] has pointed out that this assumption may overestimate the impact of disease on the whole economy and that the economic value of health reform depends on the rate of economic development. Therefore, because unemployment rates averaged approximately 7 per cent in North Karelia during the project period, income-loss estimations related to permanent disability were deflated by this figure to allow for unemployment. For short-term work losses, it was assumed that the system could not replace the temporarily disabled worker with a healthy one on short notice, so the income loss figures were not deflated.

Even with full employment, not all physically able individuals participate in the labor force. Labor force participation rates for Finland in 1975 (ages 40–44, 0.91; 45–49, 0.92; 50–54, 0.84; 55–59, 0.77; 60–64, 0.52)[13] were used to project labor force participation of the disabled had they not been disabled. Only 9 per cent of the Finnish people over the age of 64 were active in the labor force, possibly because the old age pension replaces close to 100 per cent of the person's previous earnings; it was assumed, therefore, that those over the age of 65 did not contribute to the labor force.

To be conservative in estimating hypothetical labor force participation among disabled in the absence of diability, the lowest labor force participation rate of those without CVD pensions, that of 60–64 year olds, was used, and the 0.52 labor force participation rate was applied to all CVD pensioners regardless of age.

Income Loss Estimates

The data sources for the work loss estimates are presented in Table 3.

Income Loss Due to Death. Income loss due to premature death was defined as:

$$\sum_{t=1}^{65-i} (pS_{it})(pLFP_{it})(E_{it}/1.1^t)$$

where pS_{it} is the probability that a worker of current age "i" survives for "t" more years; $pLFP_{it}$ is the probability that a worker of current age "i" will be in the labor force "t" years in the future; and $E_{it}/1.1^t$ is the annual earnings discounted annually at 10 per cent. In this analysis, "i" was set at 40, 50 and 60.

The age-stratified cause-specific number of deaths in North Karelia during the project period was available through 1975 from national statistics.[14] The numbers of deaths for 1976 and 1977 were estimated from data for 1972 through 1975. The mean age in the 35–44 age group was assumed to be 40, the mean age in the 45–54 age group was assumed to be 50, and the mean age in the 55–64 age group was assumed to be 60. The groups were assumed to have 25, 15 and five remaining working years of life, respectively, if death did not intervene because labor force participation rates after the age of 65 are less than 10 per cent. The probability of survival until the age of 65 was estimated from Finnish life tables for 1975.

Change in death rate was estimated by subtracting the change in the reference area rates from the rate changes in North Karelia:

net change in CVD death rates

$$= (NK_2 - NK_1) - (R_2 - R_1).$$

The statistical significance of the death rate change was tested using the normal

TABLE 3. Data Sources for Estimating Work Losses

Source of Work Loss	Source of Rate Estimation	Specificity of Rate Estimation
Death	Death certificate data	Age- and sex-specific rates
Permanent disability	Pension institute data	Sex-specific and age-adjusted rates
Temporary disability		
Periods lasting ≥ 8 days	Pension institute data	Crude rates
Periods lasting < 8 days	Population survey data	Crude rates

distribution:

$$Z = \frac{\text{net change in CVD death rate}}{\sqrt{p_{NK}(1 - p_{NK})(1/n_{NK_1} + 1/n_{NK_2}) + p_R(1 - p_R)(1/n_{R_1} + 1/n_{R_2})}}.$$

Income Loss Due to Permanent Disability. Although disability pensions are transfer payments, they are a good surrogate measure of disability-related income loss in the Finnish system because they are indexed to predisability income. All residents of Finland aged 15 to 64 years who are premanently unable to work because of disease may apply for a permanent pension.[15] These pensions replace approximately 45 per cent of the person's income. The national Social Insurance Institute administers all disability pensions in Finland, and a pension is awarded only after expert committee review of the patient's request and the treating physician's certification of disability. North Karelia and the reference county are in the same insurance district. Age, sex, disease and county-specific permanent pension data were available from the Social Insurance Institute of Finland for the years from 1968 to 1978 (unpublished data).

The cost of permanent disability pensions was calculated by multiplying the number of CVD-related disability pensions each year by the national average award size for that year. The annual figures were summed for the years 1972 through 1977. The earning loss due to permanent CVD-related disability was estimated by dividing the pension costs by 0.45, the reported average replacement ratio. This figure was multiplied by 0.52, the labor force participation rate in 1975 for the population aged 60–64 years, and deflated by 7 per cent to allow for unemployment effects.

Piecewise regression[16] was used to test for and estimate the size of project effects on permanent pensions. Variables for county, year (linear and quadratic components), and program (the "program" variable was constructed by assigning "0" for all years for the reference area, "1" for North Karelia in 1972, "2" for North Karelia in 1973 and so forth, through "7" for North Karelia in 1978) were regressed on total, CVD-related and non-CVD-related disability pension rates. The data were age adjusted to the 1972 population of Finland, and the analyses were performed separately for men and women. Because all CVD events occurred in persons over 35 years of age and age-specific data were available, accuracy was increased by limiting permanent pension analysis to the 35–64-year-old age group.

Income Loss Due to Temporary Disability

All working-age residents of Finland who are unable to work for periods of eight to 300 consecutive workdays are eligible for "daily allowance" payments.[17] A work year is considered to be made up of 250 workdays. Daily allowance payments pay workers 0.15 per cent of their previous year's annual income for each day that they are unable to work, providing their income falls within a defined range. If their income is below the range, it is replaced at a flat rate. If total income exceeds the range, that part of the income above the ceiling is ignored, and workers receive payment at the rate for the ceiling income. Payments to household heads are supplemented 15 per cent for a spouse and 10 per cent for each child, providing that the total supplements do not exceed 50 per cent of the basic daily allowance rate. If workers are included in the obligatory workers' insurance plan through their employers or if they receive benefits from a private insurance plan, the government reimburses the other plan for its expenses with the same formula as that used for payments to individuals. Replacement ratios are reported to average about 45 per cent of earning losses. As a rule, workers' insurance and private insurance plans do provide coverage at reimbursement levels greater than those of the Social Security Institute. However, these payments are not included in the analysis because they only represent further transfer payments and are not necessary to consider when estimating production losses.

Disease and county-specific data were available from the Social Security Institute of Finland for the years 1972 through 1977 (unpublished data). The cost to the government for daily allowance payments was calculated by summing the CVD-related daily allowance payments each year for the years 1972 through 1977. Income loss due to CVD-related disability periods greater than eight days was estimated by dividing this figure by 0.45, adding the income from eight working days for each daily allowance period initiated and then deflating by 7 per cent to allow for unemployment effects.

A difference in slopes of the pension rates in the two counties (i.e., interaction of year and county) over the five project years was defined as a sign of program effects on daily allowance costs. Dependent variables for the CVD and non-CVD daily allowance costs and daily allowance periods initiated were regressed on independent variables for county, year and county-year interaction. A p value of ≤ 0.05 was required to reject the null hypothesis of no county–year interaction.

Work losses attributable to CVD for periods totaling less than eight consecutive workdays could not be estimated directly. Therefore, they were estimated in the following manner: all survey respondents who were not pensioned, reported missing less than eight days from work or major activity in the previous year, and reported having either no disease or only CVD treated or identified by the physician in the previous year were identified. The estimations were based on this population. The work-loss rate (days/year) in subjects with no diseases was subtracted from the work-loss rate for individuals who reported having only CVD treated or diagnosed by a physician during that year. This can be described arithmetically as:

$$\text{work loss due to CVD} = (NK_{CVDO} - NK_{ND}),$$

where NK_{CVDO} is the days of work lost reported by those who missed less than eight days of work and had only CVD, and NK_{ND} is the days of work lost reported by those who missed less than eight days of work and reported no disease.

The work loss per worker was multiplied by the number of workers in the county with CVD to give the number of work-loss days in the county due to periods lasting less than eight days. The resulting number was multiplied by the calculated daily earnings to calculate the total earning losses in the country.

The calculation for the change in work loss in North Karelia over the project period then appears as:

$$(NK_{CVDO_2} - NK_{ND_2}) - (NK_{CVDO_1} - NK_{ND_1}),$$

where "2" refers to 1977 and "1" refers to 1972.

To adjust for spontaneous changes over time, the change in work loss in the reference area was subtracted from the change in work loss in North Karelia:

$$[(NK_{CVDO_2} - NK_{ND_2}) - (NK_{CVDO_1} - NK_{ND_1})] - [(R_{CVDO_2} - R_{ND_2}) - (R_{CVDO_1} - R_{ND_1})],$$

where "R" refers to the reference area.

$$Z = \frac{\text{net work loss change}}{\sqrt{S_p^2(1/n_{NK_1} + 1/n_{NK_2} + 1/n_{R_2} + 1/n_{R_1})}},$$

where

$$S_p^2 = \frac{(n_{NK_1} - 1)S^2_{NK_1} + (n_{NK_2} - 1)S^2_{NK_2} + (n_{R_1} - 1)S^2_{R_1} + (n_{R_2} - 1)S^2_{R_2}}{(n_{NK_1} + n_{NK_2} + n_{R_1} + n_{R_2}) - 4}.$$

The normal distribution was used to test for a significant change in the work-loss rate attributable to heart disease in subjects claiming less than eight days lost per year.

Results

Prevalence of CVD in the Labor Force

In 1972 and 1977, the proportion of persons on the North Karelia project surveys who claimed a CVD disease treated or diagnosed by a physician in the previous year averaged 20.5 per cent and increased from 17.8 per cent in 1972 to 23.2 per cent in 1977 in North Karelia. The proportion averaged 19.1 per cent and increased from 15.9 per cent in 1972 to 22.3 per cent in 1977 in the reference area. The net decline in North Karelia of 1.0 per cent was nonsignificant ($Z = 0.78$, $p = 0.29$).

Income Estimations

The mean income among major household providers in North Karelia was $4,031 per year in 1972 and $6,887 per year in 1977, a mean of $5,459 during the project period (all figures have been converted to U.S. dollars at a rate of 4 Fmk = U.S. $1). Assuming a work year of 250 days, these figures are equivalent to $11.04, $18.87 and $14.96 per day, respectively.

Income Loss Due to Death

From the available age-specific data, it was estimated that there were 154 CVD-related deaths among men and 39 CVD-related deaths among women in the 35–44 age group. There were an estimated 478 CVD-related deaths among men and 119 CVD-related deaths among women in the 45–54 age group and 1,084 CVD-related deaths among men and 271 CVD-related deaths among women in the 55–64 age group during the entire five-year project period. The total earnings lost to CVD-related death during the project period were $39.94 million. Although myocardial infarction death rates declined 16 per cent in men and 5 per cent in women, and stroke death rates declined 38 per cent in men and 50 per cent in women, the net changes (adjusted for changes in the reference area) were not significant.[18] Therefore, death-related income loss did not change.

Income Loss Due to Permanent Disability

In North Karelia from 1972 to 1977, government costs for CVD-related permanent disability pensions were $27.00 million. This represents income losses of $29.01 million after adjustment for income replacement ratio (45 per cent), unemployment rates (7 per cent), and labor force participation rates (52 per cent).

Among North Karelian men in 1968, about 5.5 per cent of the population aged 35–64 received a disability pension for CVD. This proportion rose to a peak of 7.9 per cent in 1974. The reference area rates started at 4.5 per cent in 1968 and continued to rise until 1976 so that the rates were similar in the two areas from 1976 to 1978 (Fig. 1). There was a significant project effect on the CVD disability pension rates ($p < 0.025$). The project effect resulted in the observed CVD pension rates averaging 6 per cent less than expected during the overall project period and 10 per cent less than expected in 1977, the last project year. Total observed pension rates tended to be less than expected, but this trend was not significant.

Among North Karelian women, 4 per cent of the population aged 35–64 was pensioned for CVD in 1968. This rate rose to a peak of 5.7 per cent in 1975. The rates in the reference community surpassed the North Karelia rates in 1971 and climbed at a steeper rate until the middle of the decade (Fig. 1). As with men, the project effect on CVD pension rates was significant ($p < 0.025$); this resulted in observed project area rates averaging 13 per cent less than expected during the entire project period and 21 per cent less than expected during the last year of the project. As with North Karelian men, observed total disability pension rates tended to be less than ex-

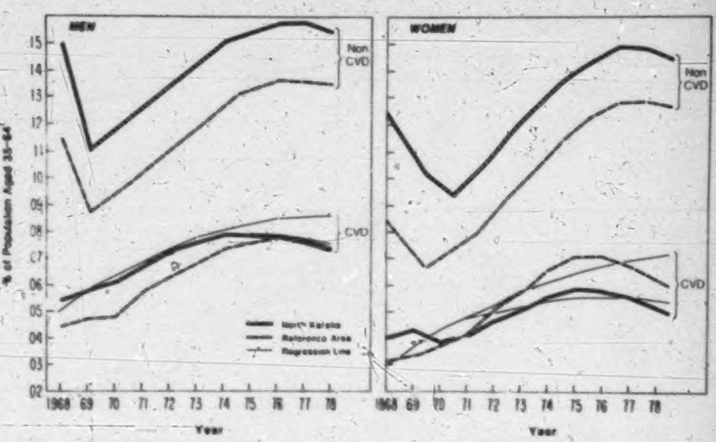


FIG. 1. Estimated by piecewise regression, the project had significant effects on both men's and women's CVD disability pension rates (both $p < 0.025$). These effects, represented by the areas between the regression lines, yielded a total savings to the government of $3.95 million during the project period. There were no project effects on non-CVD pension rates.

* Age-adjusted to 1972 Finnish population aged 35-64 years.

pected during the project period, but the trend was nonsignificant. Non-CVD pension rates were not affected by the project among either sex.

The decline in pension awards as a result of the deficit in expected CVD pensions was $1.06 million for men and $2.89 million for women—a total of $3.95 million. With adjustment for income replacement ratios, unemployment and labor force participation, this represents a $4.25 million decline in earnings loss.

Income Loss Due to Temporary Disability

Periods Greater than Eight Days Long. The total government costs for daily allowances paid in North Karelia for CVD (ICD VII) was $4.58 million from 1972 through 1977. Using a daily allowance to lost income replacement ratio of 0.45:1 and allowing for 7 per cent unemployment, this represents $9.47 million in lost earnings. Non-CVD daily allowance periods per eligible inhabitant (Fig. 2), and the CVD-related daily allowance periods per eligible inhabitant declined significantly in both North Karelia and the reference area during the project period. Although a trend was present, the rate of decline for CVD-related periods was not significantly greater for North Karelia than for the reference area ($p = 0.11$).

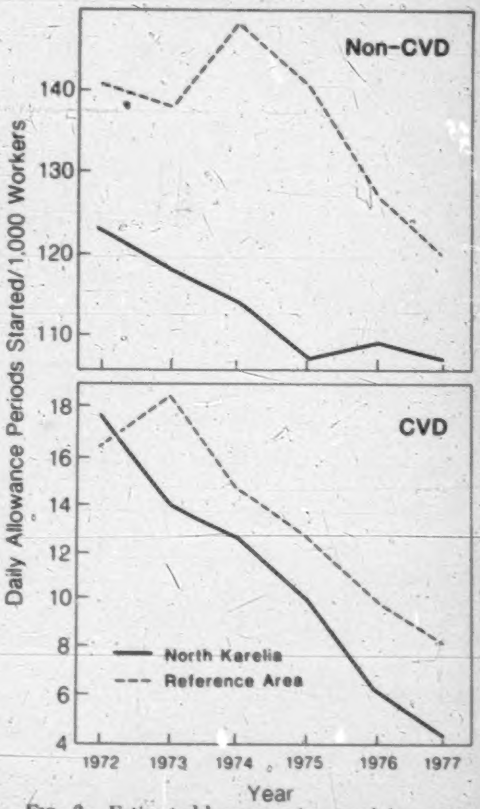


FIG. 2. Estimated by regression, work-loss costs related to temporary disability from all causes and CVD declined significantly during the project period (both $p < 0.05$). Although the decline in work-loss costs from CVD-related temporary disability was greater in North Karelia than in the reference area, this difference was not significant ($p > 0.10$).

TABLE 4. CVD-Related Temporary Disability Work-Loss Costs for North Karelia: 1972 to 1977 (million U.S.$)

	Loss*
Periods lasting >8 days	
Government reimbursed (8 to 300 workdays)	9.47
Deductible period (first 7 workdays)	0.85
Periods lasting <8 days	0.59
Total	10.91

* There were no significant changes in short-term work-loss costs.

There were 7,611 CVD-related daily allowance periods started in North Karelia from 1972 to 1977. These periods were associated with 60,888 work-loss days that were required for eligibility for daily allowance payments but not reimbursed by the system. After adjustment of unemployment, income losses for these periods were estimated to be $0.85 million.

Periods Less than Eight Days Long. In North Karelians who reported CVD and missed less than eight days from work in the year prior to the 1972 survey, work loss attributable to CVD was 1.3 days per 1,000 workdays. The 1977 figure was 0.8 day lost per 1,000 workdays. In reference area workers with CVD who missed fewer than eight days in the previous year, between 1972 and 1977 there was an increase from a negative 0.1 per 1,000 workdays to 0.4 per 1,000 workdays lost that were attributable to CVD. The resulting net decline in CVD-attributable work loss for periods less than eight working days in North Karelia of 1.0 day per 1,000 was nonsignificant ($Z = 0.035$, $p > 0.25$).

TABLE 5. CVD-Related Disability Work-Loss Costs for North Karelia: 1972 to 1977 (million U.S.$)

Loss Area	Loss	Change in Loss
Death	39.94*	Not significant
Permanent disability	29.01	−4.25 ($p < 0.025$)
Temporary disability	10.91	Not significant
Total	79.86	−4.25

* Discounted at 10 per cent to present value dollars.

Subjects claiming CVD in North Karelia during the project period averaged 20.5 per cent of the nonpensioned labor force (25,000 workers). This rate, combined with a mean work-loss rate attributable to CVD of 1.05 days per 1,000 workdays for periods of less than eight days per year in a work year of 250 workdays, gave a resulting work loss in North Karelia of 6,563 days per year, accrued in periods less than eight workdays long. This category of income loss totaled $589,095 for 1972 through 1977. Adding $589,095 in lost earnings for the illness periods of less than eight days to the income losses for periods of greater than eight days yielded an estimate of $10.91 million in income lost to temporary disability from CVD during the years 1972 through 1977 (Table 4).

Total Income Loss

Death-related income losses of $39.94 million were combined with permanent disability-related income losses of $29.01 million and $10.91 million for temporary disability-related income losses to arrive at a figure of $79.86 million in income loss due to CVD during the project period (Table 5). This amount is $4.25 million less than would have been expected without the intervention project.

Discussion

> Dollar assessments are one important measure of health benefits because, at some point in policy development, these benefits will be related to the public and private costs of bringing them about. The criterion of economic efficiency is given high priority by some policy-makers and their most influential constituencies.[19]

For this reason, the intention of this article has been to develop the truest possible estimation of CVD-related labor market losses during the North Karelia Project.

Therefore, this article does not present an estimation of total costs, and the reader should be aware that the very important contributions of those who do not directly participate in the labor force, for example, housewives, have not been counted. There has also been no attempt to account for intangible losses like pain, suffering and anxiety by assigning them a shadow price and including these estimations in the calculations.

Although it was more than four million dollars, the difference between the observed and expected income losses in North Karelia could possibly be attributed to processes other than the morbidity changes resulting from the North Karelia Project. For example, administrative decisions might have caused the observed differences. For many years, because of the social aspects of pensions, award criteria might have been lenient in economically distressed areas such as North Karelia. In the early to the mid-1970s, however, the government made an effort to standardize award criteria. Even if this effort had not been present, both North Karelia and the reference area are in the same sickness insurance district, so applications would have received similar treatment in both counties.

Another possible explanation is that migration caused the observed changes. The difference of 1,200 pensioners between the observed and expected numbers in North Karelia was only 2 per cent of the total CVD pensioners in Finland. However, it is not believed that migration explains the differences for two reasons: first, the standardized mortality rate for CVD death was only 1.48:1.00, and the ratio of pension awards for North Karelia compared with those for Finland in 1972 was 2:1. Criteria for pension awards were at least as lenient in North Karelia as anywhere else in Finland, so the pension seeker had no reason to move from North Karelia to obtain a pension. Second, the emigration from North Karelia that did occur was by young, healthy people looking for jobs in Helsinki and other cities in the south.[20] Older Finns, with their strong family attachments, would rarely leave their home area.

Economic trends could also have confounded the pension rates. In 1973, with the recession following the Arab oil embargo, Finnish exports had difficulties on foreign markets, and North Karelia suffered the most of any Finnish region. However, economic recession drives pension demand up, and the economic changes during the period would have increased, not decreased, pension demand in North Karelia.

The survey data gathered in North Karelia and the reference area also may not exactly reflect the true situation, and the methods for estimating CVD-attributable income loss for periods of less than eight days are far from perfect. Subjects have a difficult time remembering their past behavior, and responses on questionnaires about diseases and diagnoses are also subject to error. Estimates of individual income may be inaccurate because North Karelia is a rural area, and in many cases income cannot be disaggregated below the level of the family unit. Even so, because of the minor contribution of short work-loss periods to the total work-loss burden of CVD, inaccurate estimation could not have significantly influenced the results.

Conversely, to the extent that the findings of this study parallel the changes observed in morbidity, they may be interpreted as resulting from them. First, the rates for stroke declined 38 per cent for men and 50 per cent for women in North Karelia between 1972 and 1977.[18] The myocardial infarction rates declined by 16 per cent in men and 5 per cent in women in the same period. Treatment of hypertension increased equally in both counties, but the hypertensive control rate was significantly greater in North Karelia than in the reference area.[21] There was a concomitant leveling off of the population requiring drugs for congestive heart failure associated with this better control of blood pressure in North Karelía, but not in the reference area.[9] This supports the contention that the North Karelia population was relatively healthier at the end of the period than at the beginning and that the declining pension rates were due to improved health.

This study has avoided the difficulties of extrapolating future trends in disease by limiting its analysis to the five-year pilot project period. Because the changes in health behavior continue spontaneously in North Karelia at this time, however, it is believed that the declines in disability will persist at a steady rate or may even decrease further. If the differences in permanent disability were to persist in a steady state at the rate found at the end of the project, the income savings during any given period would be half again as great as those presented here. The ten-year trends will be analyzed and reported after the population is resurveyed in spring 1982.

Beyond earning losses, disability also represents opportunity costs for a community. Although the healthy worker produces while consuming and the worker who has died neither consumes nor produces, the disabled worker consumes without producing. The "Disability '72" study in the United States found that "chronic poor health impoverishes not only those it afflicts, but also those living with them"[22] because resources normally used for other purposes must be diverted to support the disabled.

Disability also causes social disorganization, and the disabled suffer from a poor self-image. Franklin[22] found that "more marriages end through divorce or separation for the disabled than for the nondisabled," and that the families of the disabled usually contract the scope of their activities to manage with the disabled person.

There are those who hesitate to examine earning losses for fear of being forced into recommending programs that focus on those with the highest per capita earnings. However, if earnings are viewed within the context of the social system, two points become apparent: first, laborers in most technological societies have gained the right of passing health care and disability costs on to their employers. It is this right that has prompted management in the U.S.[23] to seek methods of containing health care costs, and, in at least one state, management, charitable foundations and organized medicine have joined their efforts.[24] The worker who requires less health care costs management less money than the worker who requires more.

Second, although the production of any given worker may not be distinguishable from the production of another worker, the wage earner is a necessary component of technological societies. The effectiveness of the labor strike is adequate testimony to this fact. Because an efficiently working social system requires that all components be in good condition, there is little to gain by maximizing the health of the managers while neglecting the health of the wage earners.

Finally, as was done in Finland, policymakers can simply decide that preventive medicine programs must include all people, and they can adopt programs that attempt to influence the health of the whole population. Besides being equitable, it is believed that this approach may also be the most efficient public health strategy. To shun the analysis of changes in earning losses associated with public health programs during periods of social program retrenchment is to ignore some of the most powerful data supporting the efficacy and cost-effectiveness of the public health effort.

A comparison of the relative magnitudes of the project costs, the health services costs and the work-loss costs is revealing. It was previously reported that project costs were $1.02 milion for evaluation and $730,000 for implementation.[9] Although the trend analysis has not yet been completed, direct community costs (health services) totaled $13.86 million for drugs,

$10.10 million for hospital care of stroke and myocardial infarction, and $20.09 million for primary care—$44.05 million in all.

The calculations in this study represent the lower bound of CVD-related work-loss costs and only a proportion of total CVD-related costs. The work-loss costs were about 100 times the project implementation costs, and the decline in work-loss costs alone was more than six times the project implementation costs. The savings to the government in reduced pension costs alone were five times the project implementation costs. The middle or upper bound estimation would make savings in income losses even larger relative to program costs. These figures strongly suggest that the potential returns in earning capability alone merit investment in heart disease prevention programs.

References

1. Rice DP. Estimating the cost of illness. Am J Public Health 1967;57:424.

2. Cooper BS, Rice DP. The economic cost of illness revisited. Soc Secur Bull 1976;39:21.

3. Acton JP. Evaluating Public Programs to Save Lives: The Case of Heart Attacks. Santa Monica: The Rand Corporation, 1973.

4. Social Insurance Institute of Finland. Statistical Yearbook of the Social Insurance Institution of Finland. Helsinki: Social Insurance Institute of Finland, 1972. (Series T1: number 8.)

5. Black DAK, Pole JD. Priorities in biomedical research: indices of Burden. Br J Prev Soc Med 1975;29:222.

6. Smith ESO. The economic impact of preventable ischemic heart disease. Can Med Assoc J 1977;117:507.

7. Weinstein M, Stason W. Hypertension: a policy perspective. Cambridge, Mass.: Harvard, 1976.

8. Puska P, Mustaniemi H. Incidence and presentation of acute myocardial infarction in North Karelia, Finland. Acta Med Scan 1975;197:211.

9. Puska P, Tuomilehto J, Salonen J, et al. The North Karelia Project: Evaluation of a Comprehensive Community Programme for Control of Cardiovascular Diseases from 1972–1977 in North Karelia, Finland. Copenhagen: WHO Regional Office for Europe, 1981.

10. Puska P, Tuomilehto J, Salonen J, et al. Changes in coronary risk factors during comprehensive five-year community programme to control cardiovascular diseases (North Karelia Project). Br Med J 1979;2:1173.

11. Mushkin SJ, Collings Fd'A. Economic costs of disease and injury. Public Health Rep 1959;74:795.

12. Klarman HE. Economics of Health. New York: Columbia University Press, 1965;168.

13. Central Statistical Office of Finland. Statistical Yearbook of Finland—1979. Helsinki: Central Statistical Office of Finland, 1980;50-51.

14. Central Statistical Office of Finland. Causes of Death in Finland, VI B:128 to VI B:131. Helsinki: Central Statistical Office of Finland, 1975–1979.

15. Social Insurance Institute of Finland. Statistical Yearbook of the Social Insurance Institution of Finland. Helsinki: Social Insurance Institution of Finland, 1971;41, 194. (Series T1: number 7.)

16. Neter J, Wasserman W. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Homewood, Ill.: Richard D. Irwin, Inc, 1974;313.

17. Social Insurance Institute of Finland. Statistical Yearbook of the Social Insurance Institution of Finland. Helsinki: Social Insurance Institute of Finland, 1971;236. (Series T1: number 7.)

18. Salonen JT, Puska P, Mustaniemi H. Changes in morbidity and mortality during a comprehenisve community programme to control cardiovascular diseases during 1972-77 in North Karelia. Br Med J 1979;2:1178.

19. Milio N. Promoting Health Through Public Policy. Philadelphia: FA Davis, 1981;216.

20. Landers PS. In the Shadow of the Factory: Social Change in a Finnish Community. New York: John Wiley and Sons, 1976.

21. Tuomilehto J, Nissinen A, Salonen JT, Kottke TE, Puska P. Community programme for control of hypertension in North Karelia, Finland. Lancet 1980;ii:900.

22. Franklin PA. Impact of Disability on the Family Structure. Social Security Bulletin, May 1977;3.

23. Containing the cost of employee health plans. Business Week, May 30, 1977;245:74.

24. Minnesota Coalition on Health Care Costs, MC . . . health care providers, consumers, payers and educators working with business, government and labor toward cost-effective, quality health services. Minneapolis: Minnesota Coalition on Health Care Costs, undated.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Relationship of Size and Payment Mechanism to System Performance in 11 Medical Care Systems

PAUL A. NUTTING, M.D.,[*] BARTON R. BURKHALTER, PH.D.,[†]
DUANE L. DIETRICK, PH.D.,[‡] AND EDWARD F. HELMICK, MPH[§]

The performance of 11 medical care systems of varying size and payment mechanisms (consisting of six government owned and operated Indian Health Service units, three fee-for-service private practices and two HMOs) was studied. Performance was defined as the percentage of consumers in need of care who received adequate care according to predetermined standards for the process of care for various functions (i.e., prevention, screening, treatment, follow-up) and various health conditions (i.e., prenatal and infant care, hypertension, anemia, UTI). Size was found to have a strong negative relationship to the quality of treatment and follow-up care, but payment mechanism did not show such a relationship. Neither size nor payment mechanism was significantly related to prevention performance. In screening, the results depended on the health condition: size was inversely related to performance of screening for hypertension; HMOs performed significantly better in screening for prenatal anemia; and neither size nor payment mechanism was related to performance of screening for infant anemia.

DESPITE EXTENSIVE STUDY, there is not yet consensus on how the many aspects of organizational structure influence organizational performance. In the health field, much recent attention has focused on the relation of payment mechanisms to system performance, but in spite of convincing arguments that financial incentives should affect performance in health care delivery systems,[1] studies comparing fee-for-service and prepaid payment programs have produced inconclusive results on most issues,[2–9] at least in part due to methodologic problems.[9] More attention has been given to utilization and cost containment than to quality of care, and to inpatient care than to outpatient care.[8] The increased preventive care widely observed in health maintenance organizations (HMOs) is due to the prepayment mechanism that provides a low or zero marginal cost to the user, rather than to a health maintenance philosophy.[10]

Economists have investigated the relationship between firm size (an aspect of structure) and profitability in economy-of-scale studies. Some have found that larger size is generally correlated with greater profitability,[11,12] but many investigators have not found significant correlations between size and performance,[13–16]

---

[*] Associate Director for Research, Office of Research and Development, Indian Health Service, DHHS and Associate Professor (Adjunct), Department of Family and Community Medicine, University of Arizona.

[†] Senior Scientist, Community Systems Foundation and Professor (Adjunct), Department of Family and Community Medicine, University of Arizona.

[‡] Associate Professor, Department of Systems and Industrial Engineering, University of Arizona.

[§] Chief, Office of Planning and Development, Navajo Area, Indian Health Service, DHS.

All opinions expressed herein are the authors' and do not necessarily represent the policy of either the Indian Health Service or the Department of Health and Human Services.

Address for reprints: Paul A. Nutting, M.D., Office of Research and Development, Indian Health Service, P.O. Box 11340, Tucson, AZ 85734.

0025-7079/82/0700/0676/$01.25 © J. B. Lippincott Co.

and nearly all have concluded that the relationship is seriously confounded by other factors. Some investigators have concluded that firm performance is independent of structure and that firms can attain their goals by any of a wide variety of structural arrangements.[17] Studies of the relation of size to performance in community agencies have concluded that agencies of different sizes do better at different tasks.[18, 19] There is some evidence suggesting that outcomes are better in larger hospitals.[20]

Organizational complexity is another important aspect of structure that has been studied. Weber's[21] pioneering work on bureaucracy stimulated the as yet unsettled debate on the factors needed to describe organizational structure adequately, with complexity playing a central role. Some investigators have found that several dimensions are necessary to describe structure,[17, 22] with either size[19, 20] or complexity[25] being the more important dimension, while others have found that all structural factors are so highly correlated as to be inseparable.[26–30] Two large studies of government offices found that the logarithm of size was significantly correlated with indicators of complexity, such as horizontal and vertical differentiation, division of labor and personnel specialization.[29–31]

While several studies in the health field vaguely suggest that increased differentiation and specialization may detract from performance,[32–34] other studies of hospitals[35] and industrial firms[36] indicate that complexity interacts with both the coordination and the difficulty of the problems faced by the organization in determining performance. Several studies suggest that greater system complexity *from the viewpoint of the consumer* may detract from performance for high-risk consumers.[9, 37]

The study reported in this article seeks to determine how size and organizational type are related to system performance in 11 medical care systems. Performance is defined as the proportion of community members in need of care who receive adequate care according to explicit standards for the process of care. Initially, the study included six Indian Health Service (IHS) units of different size. However, the study was expanded to include three private practices and two HMOs located near the service units, in order to examine performance in three different organizational types, each type with a different payment mechanism. Thus, the study uses cross-sectional data to examine the combined effect of size and organizational type on system performance, within the limitations presented by the 11 study sites.

Study Sites

The 11 study sites include three fee-for-service private practices, two prepaid HMOs and six government-funded IHS service units. First, the six service units were selected from various locations in the western United States to represent the spectrum of service configurations in the IHS. Then, the private practices and HMOs were selected from the general geographic areas of one or more of the study service units. Specifically, private practices B and C are from the same general geographic area as IHS units C and D, while private practice A, both HMOs, and IHS units A and B are near one another.

The precise definition of the "medical care system" at each site presented problems. The "recipient population" was conveniently defined to be the prepaid subscribers at the HMOs; in the private practices, however, in which the recipient population included all past patients potentially still active in the judgment of the private practice, the definition may be inconsistent, and in the IHS units, in which the recipient population included all Native Americans known to be living within the official service area for that site, the population is not always known. In addition, the set of facilities providing service was not well delineated in any of the 11 sites. The study attempted to identify all major sources of medical service for the recipient population of each site, including major referral and "leakage" patterns. Consequently, the "system" as defined for the purposes of the study included facilities and provider groups that did not necessarily consider themselves to be components of a "system," even though patients interacted with the different components as though they were part of the system.

In no instance did a study site that was approached for inclusion in the study decline. In all cases the clinical personnel were cooperative and indicated an interest in using the study results constructively to improve their own organizations.

The typical service unit is located within a rural Native American community and serves a population of 3,000–8,000 people who may reside in smaller communities, often scattered over several thousand square miles. The typical service unit consists of a 30–50 bed hospital and outpatient department staffed by the full complement of health professionals and administrative staff. The main service unit facility serves as the primary referral center and base of operations for one or more full- or part-time field clinics, public health nurses, environmental engineers, health educators and an increasing number of problem-specific health programs (nutrition, mental health, maternal and child health, alcoholism, etc.) operated by the tribal government. Each service unit can refer patients to secondary and tertiary care centers, either operated by the IHS or through contract within the private sector.

However, two of the service units deviate substantially from this pattern. Service unit A is located on the outskirts of a large metropolitan area and consists of a single outpatient clinic staffed by a single general practitioner, a physician extender and a public health nurse. Inpatient care is provided through contractual arrangements with several large hospitals within the metropolitan area. Service unit C is located within another large metropolitan area and consists of a 250-bed medical center, a multidisciplinary outpatient clinic and two field clinics located in the surrounding rural areas. The medical center provides primary care to Native Americans living in the urban area and also serves as a referral center for several other service units, including service unit D. Service unit C is staffed by 40 physicians, of whom 32 are specialists, in addition to a physician extender and five public health nurses. The field clinics are staffed by physicians traveling from the medical center on a weekly and biweekly schedule. All of the IHS units provide primary care to Native Americans at no direct cost to the patient, as well as necessary specialty care requiring referral into the private sector.

All three of the private practices are located in rural areas and are in general proximity to one or more of the service units. Private practices B and C were the only providers of medical care in their communities. All operate on a fee-for-service basis, and participate in Medicare and Medicaid for their eligible patients. The practices consisted of two to five physicians, and one of the groups employed a physician extender. All of the private practitioners were generalists, with the exception of a pediatrician and an internist in private practice C. The physicians of all three practices had admitting privileges at small community hospitals, but none held administrative positions at the hospitals during the study time frame.

The two closed-panel HMOs are located in urban areas and in general proximity to two of the service units. One of the HMOs operated three separate clinics, and physicians in both admitted patients to two separate large urban hospitals. Both had more than 15 physicians on the full-time staff, consisting of both generalists and specialists, and both HMOs employed nurse practitioners in ambulatory care.

Many of the important structural characteristics of the 11 study sites are summarized in Table 1.

The Dependent Variable—Performance

The issues involved in defining performance of a medical care system have been extensively discussed by others.[38] A method for measuring performance that is based on explicit standards for the quality of the process of care as received by all potential consumers (not just patients) for common, ambulatory health conditions was used in the present study. This method considers neither outcomes nor costs. Since a full description of the method and its limitations has been published previously,[39] it is presented here only in summary.

A medical care system should perform various functions, including prevention, screening, treatment and follow-up, for a wide variety of health conditions. In this study, several conditions were used to examine the performance of four functions of care: prenatal and infant care was used to examine prevention performance; prenatal anemia, infant anemia and hypertension were used to examine screening performance; and anemia and urinary tract infection (UTI) were used to examine the performance of both treatment and follow-up. Explicit, minimal standards for the process of care were defined for each condition and function, and are summarized in Table 2. These standards were reviewed and agreed to as appropriate by physicians at every site.

At each study site, a sample of episodes occurring within a particular time frame during the period from 1974 through 1976 was selected for each condition, and the medical records of the persons in the sample were audited to determine if the care they received during the episode met the standard. System performance for a particular function and health condition is defined as the proportion of the entire potential recipient population in need of service that received the service according to the standard. Thus, performance is defined in terms of episodes during which an adequate process of care was received.

No single overall measure of performance for a medical care system is defined, although consistent patterns across functions and health conditions may permit overall observations on the performance of a system. Further, logical sequences of functions, such as treatment and then follow-up, can be combined to obtain an overall performance indicator for the sequence for a particular condition. Logical sequences of tasks within a particular function can also be examined, for example, *contact* between the provider and a consumer-in-need, *recognition* by the provider of the need and, finally, *provision* of appropriate care.

A prevention performance index was calculated as the average score on the six prevention tasks, while the performances for the other functions and conditions were not aggregated. General surveillance was considered part of prevention, while screening for a particular problem, such as anemia, was considered to be a separate function. It is important to note that screening does not include recognition of abnormal results. The conditions chosen are common ambulatory problems that reflect the type of situation most frequently faced by consumer and provider.

The prenatal sample used to examine prevention and screening was taken from a master list of pregnant women compiled from birth certificates, delivery room logs, hospital discharge diagnoses and requisitions for prenatal laboratory work. The master list for the infant sample was compiled from birth certificates and delivery room logs. The sample of persons between 40 and 60 years of age used to examine hypertension screening was generated from a casual selection of active medical records at each site. The master lists for UTI and anemia used to examine treatment and follow-up were obtained from laboratory logs and slip files. In order to maintain the focus on common ambulatory problems, the anemia and UTI samples were limited to simple cases by dropping episodes that showed prior complications, as defined in Table 2. Patients thus eliminated from the sample were replaced from the master list.

Sample sizes for prevention, screening and treatment generally ranged from 27 to more than 100, depending on the health condition and site, but with several exceptions.[‖] Samples tended to be smaller for the follow-up function, generally ranging from ten to 40.

Every health record known to exist for each person in all of the samples was sought out in order to capture all contacts between the person and any component of the provider system relevant to the episode in question. For the IHS units, this typically included a review of the medical record at the hospital and at one or more field clinics, public health nursing records and the records at the most frequently used referral centers. For the HMOs and private practices, records reviewed included the office records, hospital medical records and medical records at the most likely sources of referral care. In addition, an attempt was made in each study site to estimate the magnitude of "leakage" of patients outside the study site. When a major pattern was presumed to exist, the records at the major alternate sources of medical care were also audited.

This method of measuring system performance has some properties that avoid certain problems sometimes present in other methods: it samples from the entire community, not just from patients; it examines common ambulatory conditions, not just complex problems requiring hospitalization; it includes and integrates the contributions of many provider disciplines and the consumer; it examines care at many locations; and it examines the performance of several functions of care. Nevertheless, the method also has some important limitations. These include the well-known limitations of information taken from medical records, which may reflect inconsistently actual care in different patients and different systems, and the inconsistent results obtained from different measurements of quality, in which explicit process measures come out lower than other measures. While every effort was made to obtain a representative sample of all community members in need of care, the differing nature of the service populations and the aforementioned difficulty in defining service populations that were comparable across sites may have biased the results. Also, the relative success in identifying all locations at which care was received may have differed across sites. The sensitivity of the standards of care is not well understood, and small changes in the standards may produce significant changes in performance that could alter the findings. The interdependence of the functions is not well understood; for example, screening performance may influence treatment performance. More important, clear relationships between the quality of the process and the outcome have not been established. Thus, there is as yet no guarantee that better performance of process will lead to better outcomes.

The Independent Variables

Two independent variables were used in the analysis: 1) the log to the base 10 of the number of direct care providers (physi-

---

[‖] Small samples included infant prevention in HMO-B with five; infant anemia screening in HMO-B with 15; anemia and UTI treatment in IHS-A with 11 and 19, respectively; and anemia treatment in private practices A, B and C with five, nine and 18, respectively. For follow-up the anemia samples were less than ten in private practices A and B, with three and six, respectively.

cians, nurses and physician extenders), which is termed "size"; and 2) a categoric variable termed "type" that reflects whether the system is a private practice, an HMO or an IHS unit. Size is an organizational characteristic that other investigators have often found to be highly correlated with organizational complexity.[29-31] That relation appears to hold in data from this study as well, for example, in the perfect ordinal relation between size and number of disciplines, as shown in Table 1. Use of the logarithm follows previous findings and usage[29-31] and reduces the effect of outlyers.

The most apparent difference between the three organizational types is their payment mechanisms. The private practices are all fee-for-service private entrepreneurships in the traditional style of American medicine. The HMOs have prepaid subscriber plans, with a large portion of the prepayments coming from employers but with subscribers making token payments for many services. The IHS units are government owned and operated, with all provider personnel salaried or under contract. Health care is free to all eligible consumers. There are other differences among the three organizational types that might explain performance variation. Notably, the populations of American Indians and Alaskan Natives served by the IHS units are culturally distinct from the populations served by the HMOs and private practices, which were typical of the general U.S. population.

While the six IHS units vary in size from three to 46 direct care providers, the three private practices are relatively small, ranging from two to five direct care providers, and the two HMOs with 18 and 25 direct care providers are larger than all but one of the IHS units. This clustering of private practices and of HMOs on the size dimension increases the degree to which the two variables confound one another's effects on performance.

Results

Performance scores for the 11 study sites are summarized in Table 3 by function and health condition.# From these data it can be seen that, with the exception of screening for infant anemia, there is wide variation among the study sites for all functions and conditions. IHS service unit A, which is a small system with a largely suburban population, scores consistently higher than the other service units. This site had a stable and, in the authors' opinion, very able team of professionals operating the clinic, which may explain some or all of the high performance. The relative performances of the other ten systems appear to depend on the function and condition, with different systems scoring higher in one area and lower in another.

The primary goal of the analysis was to determine how size and type were related to system performance in the data. The analysis was complicated by the small sample (n = 11) and the confounding of size and type in the data. Analysis of covariance was selected as the best way to structure the analysis and to identify gross relationships, although it was clear that the data do not meet all the assumptions (particularly homoscedasticity) of the statistical model for analysis of covariance.

The combined effect of size and type on six different indicators of performance was obtained using analysis of covariance (Table 4). The null hypothesis in the test for the effect of type was, "After accounting for the effect of size, type had no effect on the variation in performance among the 11 study sites for each function and condition." This hypothesis was rejected only in

---

# The complete data, as well as details of the analysis, are available in: Nutting PA, Burkhalter BR. The Relation of Size, Organizational Type, Tracer, and Sequential Event to Performance in Prevention, Screening, Treatment and Follow-up in Eleven Medical Care Systems: A Working Paper, January 1981. DHHS, IHS, Office of Research and Development, Research and Quality Assurance Branch, Tucson, Arizona

TABLE 1. Structural Characteristics of the 11 Study Sites

Pilot Site	Approximate Population of Catchment Area	Approximate Size of Catchment Area (Sq. mi.)	Fixed Facilities	Number of Direct Care Providers				Number of Community Health Workers	Number of Direct Care[1] Disciplines
				Physicians	Physician Extenders	Public Health Nurses	Total		
IHS A	900	111	1 clinic	1	1	1	3	4	2
IHS B	6,155	2,854	50-bed hospital and OPD; 1 field clinic	5	3	5	13	10	3
IHS C	14,480	92,000	170-bed medical center and multidisciplinary OPD; 2 field clinics	40	1	5	46	35	8
IHS D	3,800	38,000	29-bed hospital and OPD	2	0	3	5	30	2
IHS E	4,926	6,375	41-bed hospital and OPD	5	1	1	7	25	3
IHS F	4,554	5,200	39-bed hospital and OPD; 1 field clinic	4	2	2	8	16	3
Private practice A	20,000	1,600	1 medical office; attending at 30-bed hospital	3	1	0	4	0	2
Private practice B	5,000	1,800	1 medical office; attending at 30-bed hospital	2	0	0	2	0	1
Private practice C	15,000	5,000	1 medical office; attending at 30-bed hospital	5	0	0	5	0	3
HMO A	22,600	2,700	2 medical offices; attending at 250-bed and 300-bed hospitals	20	5	0	25	0	6
HMO B	11,000	1,500	1 medical office; attending at two 300-bed hospitals	16	2	0	18	0	6

The number of disciplines is identified from the following predetermined list: public health nurse, physician-extender, general practitioner, internist, surgeon, pediatrician, gynecologist–obstetrician, super specialist.

TABLE 2. Standards Used to Judge the Adequacy of Care Received and Reported in the Medical Record

Function	Task or Condition	Standard
Prevention	Prenatal diet counseling	Pregnant women should receive diet counseling by 20th week of gestation
	Prenatal surveilance	Pregnant women should have fundal height measured 3 times in 2nd and 5 times in 3rd trimester, and FHR measured once in 2nd and 5 times in 3rd trimester
	Infant immunization	Infants should receive 3 DPT and 2 OPV immunizations by 13 months
	Infant diet counseling	Mothers should receive infant diet counseling 3 times before the infant is 13 months old
	Infant care counseling	Mothers should receive general infant care counseling once in 1st 6 months and once in 2nd 7 months of infant's life
	Infant growth surveillance	Infants should have weight and height measured 3 times in 1st 6 months and 2 times in 2nd 7 months of life
Screening	Prenatal anemia	Pregnant women should have a hematocrit or hemoglobin by 20th week of gestation
	Infant anemia	Infants should have a hematocrit or hemoglobin between 6 and 13 months
	Hypertension	Persons between 40 and 60 years of age should have their blood pressure measured once every 3 years (in the absence of trauma, pregnancy, intoxication or medication known to elevate blood pressure)
Treatment	Anemia	Persons screened positive for anemia (Hct < 33 or Hgb < 11, in the absence of neoplastic disease, acute or chronic blood loss, or anemia previously diagnosed as other than iron deficiency) should contact provider within 3 weeks and start iron therapy within 1 week of contact
	UTI	Persons with positive urine culture (more than 100,000 organisms, in the absence of chronic or recurrent UTI or known abnormal anatomy of urinary tract) should start appropriate antibiotic therapy within 2 weeks
Follow-up	Anemia	Persons on iron therapy for anemia should have a hematocrit or hemoglobin between 3 and 6 weeks after iron therapy was started
	UTI	Persons on antibiotic therapy for UTI should receive a urine culture or microscopic urinalysis between 1 and 4 weeks after antibiotic therapy was started

screening for infant anemia, where the significance level is 0.004.

The null hypothesis for the effect of size was, "After acconting for the effect of type, size had no effect on the variation in performance among the 11 study sites for each function and condition." This hypothesis was rejected in the treatment and follow-up sequence for both anemia and UTI, and was nearly rejected in screening for hypertension. The relationship between size and performance is strongly negative in these cases, with the smaller systems performing better than the larger ones.

These results from the analysis of covariance must be viewed as indictors of gross

TABLE 3. Performance in 11 Medical Care Systems by Functions, Health Condition, Type of Organization and Size

Function and Condition	PP			IHS						HMO		AVERAGE			
	B	A	C	A	D	E	F	B	C	B	A	PP	IHS	HMO	ALL
Prevention															
Six task average	5	19	17	56	20	21	7	26	28	26*	47	14	26	36	25
Screening															
Prenatal Anemia	49	42	51	56	48	34	33	36	43	89	71	47	42	80	50
Infant Anemia	39	30	39	36	34	31	33	32	36	33*	31	36	34	32	34
Hypertension	86	86	79	92	84	42	34	62	41	64	62	84	59	63	67
Treatment and follow-up															
Anemia	33*	40*	39*	36*	36	18	10	14	6	18	11	37	20	15	24
UTI	65	29	42	47*	43	22	16	20	8	26	18	45	26	22	31
Number of direct care providers	2	4	5	3	5	7	8	13	46	18	25				

The performance scores reflect the percentage of consumers in need of care who received adequate care for the condition and function indicated. Treatment and follow-up is reported here as a single function, with the performance scores reflecting the percentage of consumers in need of treatment who received both adequate treatment *and* adequate follow-up care. Asterisks (*) indicate a small sample (less than 25). The three organizational types are private practice (PP), Indian Health Service units (IHS) and health maintenance organizations (HMO).

relationships rather than as reliable statistical conclusions, for the reasons noted. From this viewpoint, the strong inverse relationship of size with some of the performance indicators really reflects the existence of an inverse relationship of size and performance in the data of six IHS units, with the three small private practices and two large HMOs more or less fitting into the service unit pattern.

Prevention

While neither size nor type was significantly related to prevention performance in the analysis of covariance, the *average* prevention performance of the two HMOs was considerably higher than the average performance of the three private practices or of the six IHS units. Several outlyers with especially high scores (HMO-A and IHS-A) or especially low scores (PP-B and IHS-F) contributed to this situation by increasing the variance of performance and confounding the effect of size and type even more. The differences among the three types are not consistent across the six prevention tasks. For example, the private practices scored highest in prenatal surveillance but lowest in the infant care tasks, while the IHS service units averaged highest in infant immunization and infant care counseling. Several study sites had one or more zero or near-zero performance scores, which pulled down their overall prevention performance, suggesting that some tasks were simply ignored by some of the provider systems. In fact, one service unit (IHS-C) noted that it was not performing some of the prevention tasks at the standards review and indicated its intention to change. Although the large average differences between types in prevention performance may justify further examination of the prevention function with a larger sample of medical care systems, our data certainly do not support the existence of a significant relationship.

TABLE 4. Results of an Analysis of Covariance to Find the Relation of System Type and Size to System Performance in 11 Medical Care Systems

Function and Condition	Significance of Type	Significance of Size	Direction of Relation of Size to Performance
Prevention			
Six task average	0.385	0.727	(0)
Screening			
Prenatal anemia	0.004	0.364	(−)
Infant anemia	0.648	0.942	(0)
Hypertension	0.435	0.077	(−)
Treatment and follow-up			
Anemia	0.470	0.028	(−)
UTI	0.511	0.009	(−)

Type is a categoric variable reflecting whether the system is a private practice, HMO or IHS unit. Size is the logarithm of the number of direct care providers. Performance is the proportion of consumers receiving adequate care according to explicit standards by function and condition. Treatment and follow-up is reported here as a single function. Each significance level is the probability that the variance in the performance data explained by the independent variable in question is a chance occurrence after first accounting for the variance explained by the other independent variable. The analysis of covariance used was: Subprogram ANOVA, option 7 (in Nie NH, et al. SPSS, 2nd edition. New York, McGraw-Hill, 1975). The results reported here are not necessarily statistically valid because the assumptions of the covariance model are not met, particularly homoscedasticity, and, in addition, the two independent variables (Type, Size) are seriously confounded.

Screening

Overall, approximately half of the consumers-in-need received adequate screening, but the performance differed significantly ($p < 0.001$) by health condition, with screening performance for infant anemia (34 per cent) being considerably below the performance for either prenatal anemia (50 per cent) or hypertension (67 per cent). Table 3 shows that the average screening performances are higher than either prevention performances (which were held down by numerous zero scores) or the treatment/follow-up performance. The latter reflects the longer sequence of tasks that must be performed correctly in the treatment/follow-up function and therefore allows greater possibility of error. In fact, treatment performances alone are generally higher than screening performances.

The effect of size and type on screening performance also differed by health condition. For infant anemia, neither size nor type was important; for prenatal anemia, type was significantly related to performance (at the 0.004 level), but size was not; and for hypertension, size was related to performance (at the 0.077 level), but type was not. For prenatal anemia, the average performance of the HMOs (80 per cent) was considerably higher than the average performance of either the private practices (47 per cent) or the IHS units (42 per cent).

Screening requires the successful completion of two events in sequence—the first is *contact* between the consumer and provider, and the second is *provision* of an appropriate screening test once contact has been made. Provision of screening once contact had been made was performed adequately in about the same proportion of consumers-in-need in all three health conditions (averaging 66 per cent in infant anemia, 74 per cent in prenatal anemia and 75 per cent in hypertension), while contact rates differed considerably among the health conditions (averaging 53 per cent in infant anemia, 74 per cent in prenatal anemia and 87 per cent in hypertension).

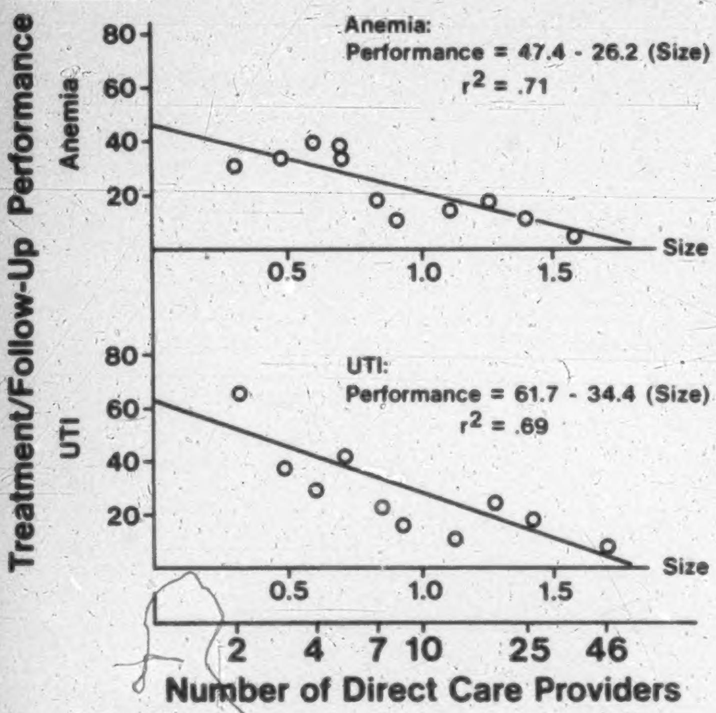


FIG. 1. Linear regression of size with treatment /follow-up performance for anemia and UTI in 11 medical care systems, where size is the logarithm of the number of direct care providers in the system and performance is the proportion (out of 100) of consumers in need of care who received adequate treatment and follow-up care. This figure demonstrates the increasing performance in curative care in smaller systems.

Contact rather than provision was responsible for the superior performance of the HMOs in relation to the other types in prenatal anemia screening. However, in hypertension screening, the inverse relation of size to performance appears primarily in provision rather than contact.

Treatment and Follow-up

Treatment and follow-up have been combined into a single sequence of care and one measure of performance, used for both functions, that reflects the proportion of consumers in need of treatment who received both adequate treatment and adequate follow-up care. This was done 1) because treatment and follow-up are a logical sequence, and the same sample of episodes was used to explore both functions in each study site, and 2) because the sample sizes for follow-up are rather small and biased in favor of individuals who receive adequate treatment in that they contain the subset of the treatment sample that received adequate treatment.

The results are dramatic. Size shows a strong, significant and negative correlation with performance of treatment/follow-up while type is not related. Small systems do better, as seen in Tables 3 and 4, and in Figure 1. The analysis of covariance produced significance levels for the effect of size on performance, after accounting for type, of 0.028 for anemia and 0.009 for UTI, while both regression lines in Figure 1 differ from zero at a significance level less than 0.005.

The inverse relationship of size to treatment/follow-up performance is not only statistically significant; it is also large. The smaller systems (2–5 direct providers) experienced a 30–50 per cent performance in treatment/follow-up, compared with 10–25 per cent in the larger systems (seven or more direct care providers); in other words, the smaller systems were roughly twice as successful as the larger ones. Within the range of the present data, when the number of direct care providers doubled, the performance was reduced by approximately another ten percentage points.

The treatment and follow-up function requires successful completion of six events in sequence: *contact* between the consumer and provider for treatment; *recognition* by the provider that treatment is needed; and then *provision* of the needed treatment, followed by contact, recognition and provision of follow-up care. In order to gain more insight into the strong and significant inverse relation between size and the performance of the treatment/follow-up function, the performance of various components of treatment and follow-up was examined, including the relation of size to treatment only; to follow-up only; and to contact, recognition and provision of service. While average treatment performance (72 per cent) was superior to follow-up performance (41 per cent), as shown in Table 5, the significant inverse relation between size and performance held for both treatment and follow-up. When size was regressed with performance of anemia treatment, with performance of UTI treatment, with performance of anemia follow-up and with performance of UTI follow-up, the slope was negative and differed from zero at the 0/05 significance level in all four cases.

In addition, size was regressed against all contact performances (anemia treatment contact, UTI treatment contact, anemia follow-up contact and UTI follow-up contact) for each of the 11 study sites, and also with all recognition and provision performance scores. The results are shown in Table 5. Average provision performance (88 per cent) is superior to the average performance of either contact (77 per cent) or recognition (80 per cent). The regression lines are negative in slope for all three sequential events, but significantly so only for recognition and provision. IHS units with larger catchment areas have more portals of entry in the form of more field clinics and community health workers than the other study sites, a strategy that

TABLE 5. The Results of Regression Analysis Relating Size to Performance of Contact, Recognition, and Provision in Treatment and Follow-up for Anemia and UTI, and Comparing Average Treatment and Follow-up Performances

Regression Equation	r^2 *	Probability that Slope = 0	Average Treatment Performance	Average Follow-up Performance	Overall Average Performance
Contact performance = 83.9– 8.1 (Size)	0.05	p = large	85%	68%	77%
Recognition performance = 97.0–24.4 (Size)	0.32	p = 0.001	84%	66%	80%
Provision performance = 100.6–13.6 (Size)	0.31	p = 0.001	90%	87%	88%
Percent of consumers with adequate contact, recognition and provision			72%	41%	

This table demonstrates the negative and, for recognition and provision, statistically significant relationship between size and performance of the sequential events of contact, recognition and provision in the treatment and follow-up functions for anemia and UTI. Treatment performance is superior to follow-up performance. Each regression equation used 44 observations, with each regression using the following four observations at each of the 11 sites: anemia treatment, anemia follow-up, UTI treatment and UTI follow-up.

may be responsible for maintaining their contact performance roughly equal to that of the geographically smaller systems.

Thus, the inverse relation between size and performance during treatment and follow-up is pervasive in these data, being found in both anemia and UTI, in both treatment and follow-up independently, and in recognition and provision, although the samples used to examine follow-up as an independent function are small and biased, as noted.

Discussion

Eleven medical care systems of varying size and payment mechanisms (consisting of six government funded and operated IHS units, three fee-for-service private practices and two prepaid HMOs) were analyzed with respect to their performance for various functions (prevention, screening, treatment and follow-up) and health conditions (prenatal and infant care, hypertension, anemia and UTI). Size was defined as the logarithm of the number of direct care providers in each system, while performance was the percentage of potential consumers in need of care who received adequate care according to explicit standards for the process of care.

Size was found to have a strong, significant and negative correlation with performance of both the treatment and follow-up functions, but payment mechanism was not. Within treatment and follow-up the strong inverse relation was also true for recognition and provision, but not for contact. In screening, the relationship of size and payment mechanism to performance depended on the health condition: prepaid HMOs performed significantly better in screening for prenatal anemia, but size was not a significant factor; small systems performed better than large systems in screening for hypertension, but payment mechanism was not a significant factor; neither size nor payment mechanism influenced screening for infant anemia. In prevention, neither size nor payment mechanism was statistically important in explaining performance, although the HMOs had a higher average prevention performance than the service units or private practices. There was no evidence to contradict the previously reported finding[10] that HMOs provide better preventive care, because the marginal cost of preventive care to the consumer is very low.

These results tend to support previous findings that organizations of different size do relatively better at different tasks.[18,19] In the present case, smaller organizations tended to do relatively better at curing tasks than at screening or prevention. Smaller systems were also better at recognition, which fits previous findings that continuity of provider increases recognition performance[34] if one assumes that patients are less likely to encounter the same provider in a larger system. The higher provision performance observed in smaller systems may also be due to greater continuity of provider. In larger systems, there may be not only a greater liklihood that patients will contact a different provider in a different location from one visit to the next, but also a greater liklihood that the provider who is contacted will not be able to perform the needed service, even though he/she recognizes the need for it. For example, a public health nurse is not trained to prescribe certain treatments, and a field clinic may not have the facilities for certain procedures. Of course, larger systems can organize themselves in various ways to try to maximize continuity of provider and minimize wasted encounters, although systematic efforts by the larger systems in this study to do that were not evident. The fact that contact performance did not vary significantly with size suggests that the multiple portals of entry in the large IHS systems have been effective in increasing access to care.

The data from this study also tend to support previous findings of high correlations between size and complexity, although the finding in this regard is based on very little data. The data also support previous findings[29–31] that the logarithm of size relates more strongly to complexity than does size alone.

With the exception of screening for prenatal anemia, organizational type does not appear to be a very important predictor of performance, and, thus, the differences in payment mechanisms represented here do not appear to be important.

However, these conclusions must be viewed within the limitations of the study, and certain caveats must be considered before generalizing the results to other settings. First, although there was no apparent bias due to self-selection (no site approached for the study declined participation), the study sites were not chosen randomly and are not necessarily a representative cross section of their organizational types. Second, no urban or large multispecialty private practices were included in the study, and no inference should be made that the three rural private practices represent the much larger universe of fee-for-service medical care. Third, the number of sites studied was small (only 11), and in some cases the number of observations for particular functions and health conditions were also quite small. The confounding of size with type of payment mechanism is a further limitation resulting partially from the small sample (for example, the sample contained only small private practices and only relatively large HMOs). Fourth, the study examined the performance of the medical care system only for basic issues in medical care. The strongly negative correlations observed in treatment and follow-up might not have been present if the study had focused on more complex medical problems. Fifth, the performance measure itself has a number of limitations, as summarized above. Sixth, numerous other variables that have not been systematically considered here might better explain the observed variation in performance. One of these, management capability, was observed informally and, except for the case noted above, did not appear to vary noticeably across study sites. Other aspects of organizational complexity, such as the relative use of physician extenders and public health nurses in place of physicians, may also be important. Finally, of course, this is a study of correlations among cross-sectional data, with no claim to making causal inferences based on appropriate controls.

In light of these limitations, the results should be the cause for hypotheses rather than the cause for action. Do the results hold for a larger and better controlled sample of medical care systems? Do they hold for a wider variety of health conditions? Does the apparent inability of the larger systems in this study to provide adequate treatment and follow-up care reflect simply a lack of management technology for control and coordination, as suggested by some investigators,[35,36] or does it reflect faults in basic assumptions underlying the medical care delivery system, as suggested by others?[40–42] Can the suspected deleterious effects of organizational size be corrected by organizational arrangements, such as subgroups? Would the introduction of costs into the definition of performance alter the results? In any case, the results of this study may have important implications for those who advocate increasingly large and complex health care systems to meet the growing demands for the quality and quantity of medical services.

References

1. Pauly MV. Efficiency, incentives, and reimbursement for health care. Inquiry 1970;7:114.

2. Dutton DB. Patterns of ambulatory health care in five different delivery systems. Med Care 1979;17:221.

3. Roemer M, Shonick W. HMO performance: the recent experience. Milbank Mem Fund Q 1973;51:271.

4. Slesinger D, Tessler R, Mechanic D. The effects of social characteristics on the utilization of preventive medical services in contrasting health care programs. Med Care 1976;14:392.

5. Gaus DR, Cooper BS, Hirschman CG. Contrasts in HMO and FFS performance. Soc Security Bull 1976;39:5.

6. LoGerfo JP, Larson E, Richardson WC. Assessing the quality of care for urinary tract infection in office practice: a comparative organizational study. Med Care 1978;16:488.

7. Shortell S, Richardson W, LoGerfo JP, et al. The relationships among dimensions of health services in two provider systems: a causal model approach. J Health Soc Behav 1977;18:139.

8. Richardson WC, Diehr PK, LoGerfo JP, et al. Comparisons of prepaid health care plans in a competitive market: the Seattle prepaid health care project. DHEW Publ No (PHS) 80-3199. NCHSR Research Summary Series, August 1980.

9. Wolinsky FD. The performance of health maintenance organizations: an analytic review. Milbank Mem Fund Q 1980;58:537.

10. Luft HS. Why do HMOs seem to provide more health maintenance services? Milbank Mem Fund Q 1978;56:140.

11. Miller EM. The extent of economies of scale—the effects of firm size on labor productivity and wage rates. Southern Econ J 1978;44:470.

12. Rakowski JP. Cost differences according to firm size in U.S. trucking. Transportation J 1980;18(2):63.

13. Singh A, Whittington G. Growth, Profitability and Valuation. Cambridge: Cambridge University Press, 1968.

14. Bethlehem G. Note on growth, profitability and size of firms. S Afr J Econ 1979;47:198.

15. Soete LLG. Firm size and inventive activity—evidence reconsidered. Eur Econ Rev 1979;12:319.

16. Buckley PJ. Influence of firm size, industry, nationality, and degree of multinationality on growth and profitability of world's largest firms. Weltwir Arc 1978;114:243.

17. Reimann BC. On the dimensions of bureaucratic structure: an empirical reappraisal. Adm Sci Q 1973;18:462.

18. Ostrom E, Parks RB, Whitaker GP. Some evidence of effects of policy agency size. The Police Chief 1978;45(2):28.

19. Kovack KA. Organization Size, Job Satisfaction, Absenteeism and Turnover. Washington, D.C.: University Press of America, 1977. (Reviewed by Muchinsky PM in Personnel Psych 1978;31:436.)

20. Scott Wr, Flood AB, Ewy W. Organizational determinants of services, quality and cost of care in hospitals. Milbank Mem Fund Q, 1979;57.

21. Weber M. The Theory of Social and Economic Organization. Translated by Henderson AM. Parson T. New York: The Free Press, 1946.

22. Pugh DS, Hickson DJ, Hingings CR, Turner C. Dimensions of organization structure. Adm Sci Q 1968;13:65.

23. Shortell SM, Getzen TE. Measuring hospital medical staff organizational structure. Health Sys Res 1979;14(2):97.

24. Terrien FC, Mills DC. The effect of changing size upon the internal structure of an organization. Am Sociol Rev 1955;20:11.

25. Anderson TR, Warkov S. Organizational size and functional complexity. Am Sociol Rev 1961;26:23.

26. Hall RK. Intraorganizational structural variation: application of the bureaucratic model. Adm Sci Q 1962;7:295.

27. Blau PM. The hierarchy of authority in organizations. Am J Sociol 1967;23:453.

28. Child J. Organization structure and strategies of control: a replication of the Aston studies. Adm Sci Q 1972;17:163.

29. Beyer JM, Trice HM. A reexamination of the relations between size and various components of organizational complexity. Adm Sci Q 1979;240:48.

30. Blau PM. A formal theory of differentiation in organizations. Am Sociol Rev 1970;35:201.

31. Blau PM, Schoenherr RA. The Structure of Organizations. New York: Basic Books, 1971.

32. Morehead MA, Donaldson R. Quality of clinical management of disease in comprehensive neighborhood health centers. Med Care 1974;12:302.

33. Morehead MA. The medical audit as an operational tool. Am J Public Health 1967;57:1643.

34. Starfield BH, Simborg DW, Horn SD, Yourtee SA. Continuity and coordination in primary care: their achievement and utility. Med Care 1976;14:625.

35. Longest B. Relationships between coordination, efficiency and quality of care in general hospital. Hosp Adm 1974;19(14):65.

36. Lawrence PR, Lorsch JW. Differentiation and integration in complex organizations. Adm Sci Q. 1967; 12:1.

37. Nutting PA, Barrick JE, Logue SC. The impact of a maternal and child health care program on the quality of prenatal care: an analysis by risk group. J Community Health 1979;4:267.

38. Donabedian A. Needed Research in the Assessment and Monitoring of the Quality of Medical Care. National Center for Health Services Research Report, July 1978. (DHEW Publication No. (PHS) 78-3219.)

39. Nutting PA, Shorr GI, Burkhalter BR. Assessing the performance of medical care systems: a method and its application. Med Care 1981;19:281.

40. Navarro I. The underdevelopment of health of working America: causes, consequences, and possible solutions. Am J Public Health 1976;66:538.

41. Illich I. Medical Nemesis: The Expropriation of Health. London: Calder and Bayars, 1975.

42. Carlson R. The End of Medicine. New York: John Wiley and Son, 1975.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Effect of a WIC Program on Children's Clinic Activity in a Local Health Department

JONATHAN B. KOTCH, M.D., M.P.H.,* AND DAVID WHITEMAN, M.A., PH.D.†

Two years' encounter data from a local health depratment in rural North Carolina are analyzed to determine the extent to which WIC increased the demand for health department services. WIC clinic visits were found to contribute disproportionately to Children's Clinic activity. Nearly 80 per cent of the increase in total clinic encounters from FY78 to FY79 was attributable to WIC alone. Whereas in the first quarter of FY78 only 18 per cent of Children's Clinic encounters by infants were to WIC clinics, that number had risen to more than 58 per cent by the last quarter of FY79. The number of children on WIC as a percentage of all non-WIC encounters with public health nurses increased during the same period from 19 to 34 per cent. Finally, infants who came for the first time to the Children's Clinic to be certified for WIC contributed an increasing percentage of all infants seen in non-WIC clinics, from 0 per cent at the start of the study to almost 15 per cent by the final quarter. It is concluded that the presence of the WIC program in this local health department was responsible for substantial increases in non-WIC clinical activity that were not supported financially by WIC or any other source.

THE SPECIAL SUPPLEMENTAL FOOD PROGRAM for Women, Infants, and Children (WIC) is now one of the "most effective and successful" child nutrition programs in the country.[1] Throughout the formal consideration of this program, which now spans more than a decade of debate, demonstration and nationwide implementation, its proponents have based their support on the significant long-term health benefits resulting from food supplements for pregnant women, breast-feeding mothers, and infants and children under five years of age who are at nutritional risk. On the strength of this argument, the WIC program has grown from a $20-million pilot study in 1972 to a $750-million program in 1979.

Most evaluation studies of WIC have focused on the physiologic and anthropometric effects of nutrition supplementation on mothers and children.[2,3] The focus of this article, however, is the consequences of the decision to place WIC programs in health agencies. Two arguments have been advanced in support of a health locus for WIC. One argument is that such placement would "encourage low-income families to come to the health centers so that total maternal and child care services can be provided."[4] A second argument can be traced to the urgings of Senator Hubert Humphrey, author of the original legislation, who said that such placement would ensure that the

---

* Department of Maternal and Child Health, School of Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

† Department of Government and International Studies, University of South Carolina, Columbia, South Carolina.

Supported in part by the North Carolina Area Health Education Center (AHEC) Program. Dr. Whiteman was supported by a research assistantship provided by the Department of Maternal and Child Health, School of Public Health, The University of North Carolina at Chapel Hill.

Address for reprints: Jonathan B. Kotch, M.D., M.P.H., Department of Maternal and Child Health, School of Public Health, The University of North Carolina at Chapel Hill, Rosenau Hall 201-H, Chapel Hill, NC 27514.

0025-7079/82/0700/0691/$00.90 © J. B. Lippincott Co.

program would "not be mismanaged in terms of some of the other programs."[5] Recent Congressional debate has made more clear the desire of WIC supporters that the program be perceived as a health, rather than a welfare, program.[6]

However, throughout Congressional consideration of the WIC program, supporters have failed to anticipate the additional demands that WIC would make on already overcrowded health department services, particularly on those provided by public health nurses. Similarly, past evaluators of the WIC program have failed to take this into consideration. This article attempts to address this deficiency by exploring the experience of one public health department in a rural North Carolina county.

Background

The population of Columbus County, in southeastern North Carolina, was 91 per cent rural in 1970. The per capita income, $4,349 in 1977, ranked 87th among the state's 100 counties. The infant mortality rate, 23.3 per 1,000 live births in 1977, is higher than that of the state as a whole, which in turn is higher than that of the nation. There are no pediatricians in the county, and its physician/population ratio, 2.4 per 10,000 in 1975, puts it in the bottom fifth of the state.

In late 1974, the County Health Department opened a Children's Clinic, utilizing the services of its newly trained pediatric nurse practitioner, a public health nurse, an outreach worker and a clerk. In March 1977, Columbus was one of four counties that undertook jointly a WIC program. With the arrival of the county's WIC nutritionist in November 1977, the program's operation began in earnest.

Methods

The conclusions of this study are based largely on data collected through June 30, 1979, the end of the second full year that a computerized, encounter-based information system was in operation in the Children's Clinic.‡ Note that, except for the special analysis described below, the unit of measurement is the encounter, not the child. For the purpose of assessing changes in Clinic activity, the number of encounters is more appropriate than an unduplicated count of children. During these two years, the Children's Clinic recorded a total of 5,983 encounters with infants, children and youths aged 0–21 years.

Not all Health Department encounters with children occur in the Children's Clinic, so the information system does not capture the universe of all child visits. For example, infants and children presenting only for immunization do not go through the Children's Clinic. All child visits for health screening, health maintenance, acute illness and follow-up care occur in the Children's Clinic.

Encounters have been classified into three types: WIC, well baby/well child ("well"), and all other encounters ("other"). The WIC category includes clients who come to the clinic for semiannual WIC certification or recertification, but not clients who come for WIC nutritional counseling. Such counseling visits with the WIC nutritionist, which do not entail a clinical service, are not included as encounters. The "other" category includes such services as episodic visits, follow-ups laboratory screening only and referrals from the nurse to the physician.

A special subanalysis was undertaken to track individual children who first appear at a WIC clinic but who later visit a non-WIC clinic. In order to capture that population of children whose first visit to any Health Department clinic occurred during the period of study, only children born on or after July 1, 1977, are included in the subanalysis. Of the total of 5,983 Children's Clinic encounters recorded between July 1, 1977, and June 30, 1979, 2,946 were accounted for by this group of 1,258 infants.

---

‡ Since September 1976, activity in the Columbus County Children's Clinic has been monitored by the Maternal and Child Health Department of the University of North Carolina School of Public Health. The Department has also recruited physicians who, while Master's candidates in the school, provide pediatric backup for the clinic's nurses. After several months of trial operation, a computerized information system was fully implemented on July 1, 1977. The system provides information of every visit to the Children's Clinic, including encounter type, provider type, health problem(s), referral(s), screening test result(s), and demographic and anthropometric data.

Results

The assessment of WIC impact on Children's Clinic activity can be separated into two parts. First, the general impact of the program on total clinic activity can be ascertained by examining the frequency of all clinic encounters, including WIC visits. Second, the more specific impact of WIC on the activities of public health personnel can be analyzed by examining the frequency of encounters processed by the Clinic's nursing staff.

In the first case, total activity in the Clinic increased substantially during the two-year period. Total Clinic encounters increased 32.5 per cent from FY78 to FY79. Figure 1 displays quarterly data divided into the three encounter categories. The rate of increase for WIC encounters is much greater than the rate for the other two categories. According to Table 1, while the "well" encounters increased by 19 per cent and the "other" encounters increased by five per cent, the number of WIC encounters grew from 1,241 to 1,910, an increase of 54 per cent. The nursing staff saw 2,240, or 71 per cent, of these 3,151 encounters, 926 in FY78 and 1,314 in FY79.

Table 1 also reveals some seasonal oscillation in Clinic activity. By comparing the total encounters for each quarter of the second year with those of the same quarter of the first year, such seasonal variation can be controlled. The resulting figures indicate that the increase in WIC encounters in all cases made the largest single contribution to the total increase between each pair of quarters. "Well" and "other" encounters also contributed to the overall increase, but to a lesser extent.

In order to examine further the demand that WIC clients made on activity in well-child and other clinics (i.e., "non-WIC" clinics), a subanalysis of infants born on or after the initiation of data collection was undertaken. This analysis shows that the proportion of all visits made by this group to a WIC clinic increased from 37.7 per cent in the first fiscal year to 62 per cent in the second. Similarly, when one looks only at first visits made by these infants, 36.4 per

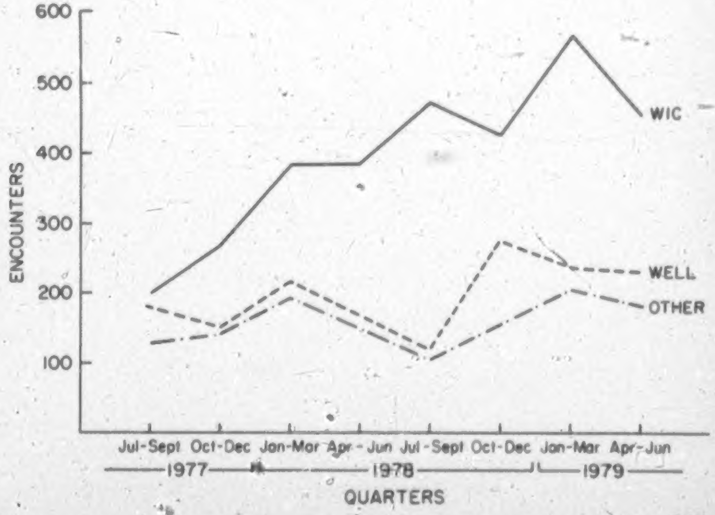


FIG. 1. Total clinic encounters by type of encounter, Columbus County Children's Clinic, July 1, 1977, through June 30, 1977 by quarter year.

TABLE 1. All Encounters per Quarter and per Year, by Encounter Type, Columbus County Children's Clinic, FY78 and FY79

	Quarter Years				Total		
Encounter Type	July–Sept.	Oct.–Dec.	Jan.–March	April–June	Annual Number	(Per Cent Increase)	Per Cent Contribution to Total Annual Increase
WIC							
FY79	471	425	563	451	1910		
FY78	199	271	385	386	1241		
Increase	272	154	178	65	669	(53.9)	79.9
Well							
FY79	117	273	234	227	851		
FY78	178	150	215	171	714		
Increase	−61	123	19	56	137	(19.2)	16.4
Other							
FY79	104	160	202	183	649		
FY78	128	143	194	153	618		
Increase	−24	17	8	30	31	(5.0)	3.7
Total							
FY79	692	858	999	861	3410		
FY78	505	564	794	710	2573		
Increase	187	294	205	151	837	(32.5)	100

cent in the first year were to a WIC clinic, compared with 58.6 per cent in the second. Table 2 displays the encounter data for these infants for each individual quarter year.

The extent to which those infants who initially came to the Health Department to be certified for WIC but who returned for a non-WIC clinic was addressed by looking at the unduplicated counts of such infants as a proportion of infants visiting all children's clinics and as a proportion of those infants visiting only non-WIC clinics (Fig. 2). In both cases, the proportions increased. In the first quarter, July 1 through September 30, 1977, no infants visited a WIC clinic before visiting a non-WIC clinic. However, the percentage of infants seen in any encounter in the Children's Clinic whose first visit ever was to a WIC clinic rose from 2.10 per cent in the second quarter to 5.39 per cent in the last quarter. An even steeper increase was observed when the unduplicated number of infants visiting non-WIC clinics only is examined. The percentage of such infants who first visited a WIC clinic rose from 3.19 per cent in the second quarter to 14.81 per cent in the final quarter (Table 3).

The influence of the expanding WIC Program on the work load of existing public health personnel can be assessed by examining encounters for all children with the Clinic's two full-time nurses. The total number of all categories of encounters with the nursing staff increased 33.6 per cent during the study period, a slightly greater percentage increase than that for all Clinic providers combined. The addition, in November 1977, of a full-time WIC nutritionist did not mitigate the new demands made on nurses' time.

Nursing staff encounter figures for similar quarters are compared in Table 4. The results differ from the results of all encounters (Table 1) in that, in one pair of matched quarters, the "well" category is the largest component of the total increase in nursing

TABLE 2. Visits to Columbus County WIC Clinics by Infants Born on or After July 1, 1977, by Quarter Year

	All Visits			First Visits		
		WIC Clinics			WIC Clinics	
	All Clinics	Number	Per Cent	All Clinics	Number	Per Cent
1977						
July–Sept.	154	28	18.2	125	27	21.6
Oct.–Dec.	176	54	32.4	102	43	42.2
1978						
Jan.–March	306	136	44.4	160	74	46.3
April–June	306	134	43.8	132	45	34.1
July–Sept.	447	295	66.0	215	148	68.8
Oct.–Dec.	442	280	63.4	175	104	59.4
1979						
Jan.–March	621	380	61.2	195	99	50.3
April–June	494	288	58.3	152	82	54.0

staff activity. However, the WIC increase is the largest component in the other three cases and consequently accounts for the largest proportion of the total annual increase. WIC was also the only category of encounters to show an increase in each of the four quarters, including the first quarter when total encounters went down.

One final indicator of WIC's influence on nursing staff activity, the proportion of

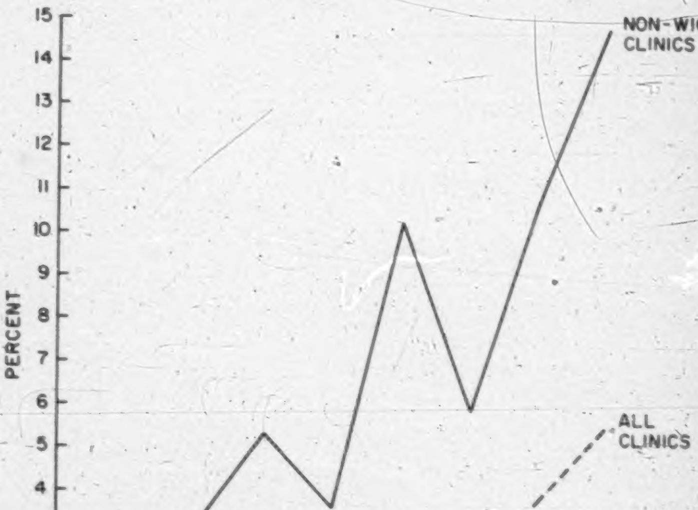


FIG. 2. Unduplicated infants whose first visit to Columbus County Children's Clinic was to a WIC clinic, as a percentage of all infants born on or after July 1, 1977, who were seen in all clinics or in non-WIC clinics, by quarter year.

TABLE 3. Unduplicated Counts of All Infants Born on or After July 1, 1977, and of Such Infants Whose First Visit to Columbus County Children's Clinic Was to a WIC Clinic, Seen in All Clinics and in Non-WIC Clinics, by Quarter Year

	All Clinics			Non-WIC Clinics Only		
		Infants First Seen for WIC			Infants First Seen for WIC	
	Total Infants	Number	Per Cent	Total Infants	Number	Per Cent
1977						
July–Sept.	125	0	0	98	0	0
Oct.–Dec.	143	3	2.10	94	3	3.19
1978						
Jan.–March	251	7	2.79	130	7	5.38
April–June	258	5	1.94	138	5	3.62
July–Sept.	407	13	3.19	127	13	10.24
Oct.–Dec.	395	8	2.03	135	8	5.93
1979						
Jan.–March	526	20	3.80	186	20	10.75
April–June	430	24	5.58	162	24	14.81

"well" and "other" encounters made by clients already on WIC, is shown on Table 5. The data, available only after April 1978, demonstrate a steady increase. During the 1978 quarter from April to June, clients on WIC accounted for 19 per cent of non-WIC encounters. By the final quarter, such clients accounted for 34 per cent of non-

TABLE 4. Nursing Encounters per Quarter and per Year, by Encounter Type, Columbus County Children's Clinic, FY78 and FY79

	Quarter Years				Total		
Encounter Type	July–Sept.	Oct.–Dec.	Jan.–March	April–June	Annual Number	(Per Cent Increase)	Per Cent Contribution to Total Annual Increase
WIC							
FY79	199	304	457	354	1314		
FY78	148	210	313	255	926		
Increase	51	94	144	99	388	(41.9)	62.6
Well							
FY79	87	260	198	198	743		
FY78	159	144	185	124	609		
Increase	−69	116	13	74	134	(22.0)	21.6
Other							
FY79	56	95	135	122	408		
FY78	69	48	104	89	310		
Increase	−13	47	31	33	98	(31.6)	15.8
Total							
FY79	342	659	790	674	2465		
FY78	373	402	602	468	1845		
Increase	−31	257	188	206	620	(33.6)	100

WIC encounters. Therefore, not only did the "well" and "other" nursing encounters increase substantially (22 and 32 per cent, respectively) from one year to the next, but the proportion of encounters attributable to children on WIC seen in those clinics increased steadily.

TABLE 5. Non-WIC Nursing Encounters by Clients on WIC

	Total Non-WIC Encounters	Number With Clients on WIC	Per Cent
April–June 1978	213	40	19%
July–Sept. 1978	143	33	23
Oct.–Dec. 1978	355	86	24
Jan.–March 1979	333	85	26
April–June 1979	320	108	34

Discussion

Congress passed the WIC Program with the dual intent of conferring the health benefits of food supplements on pregnant women, breast-feeding mothers, and infants and children at risk, and of improving the overall health status of program participants. The legislative history of the program indicates that, in placing the WIC Program in health agencies, sponsors hoped to avoid the onus of a "welfare program" label. Little serious thought was given to the consequences of attracting a new constituency to local health departments. In particular, no provision was made for the increased demand WIC clients would make on those health department staff also providing non-WIC services.

It has been demonstrated that the total encounters in the Columbus County Children's Clinic increased remarkably during a two-year period, coinciding largely, but not completely, with the growth of the WIC program. There were 32.5 per cent more total encounters in FY79 than in FY78. Total WIC encounters increased 54 per cent during the same period. This is consistent with an increase in the average monthly active case load of WIC infants and children of 52 per cent over the same period.§

The attempt to substantiate the impression that the WIC program was contributing to increasing health department work load in other clinic areas led to a separate analysis of encounter data for infants who were born since the data system was implemented. This is the group of those children who would not have attended a Health Department clinic prior to July 1, 1977. The data indicate not only that WIC clinic visits account for an increasing proportion of all encounters for these infants, but that more and more infants who first visit the health department for WIC certification later appear for a well-child or other clinic encounter. While it is impossible to prove that these infants would not have visited the Children's Clinic anyway, all of the encounter data strongly suggest that these infants and children who come to the Health Department for WIC certification contribute disproportionately to the increasing activity in all other aspects of Columbus County Children's Clinic operations.

In order to evaluate further the findings that WIC may account for much, if not most, of the increased demand on the Children's Clinic, Columbus County information was compared with information from two North Carolina counties without WIC programs. Counties without WIC were chosen for comparison according to similarities in six criteria: geographic location, total population, infant mortality rate, percentage under five years of age, percentage nonwhite and percentage of AFDC.‖

---

§ The active case load is the total number of infants and children whose WIC coupons were collected from the Health Department in a given month. Data are from the WIC program of the Columbus County Health Department.

‖ The source of this information is the North Carolina Department of Administration, Division of Budget and Management.

The two counties chosen, Beaufort and Duplin, were found to have had increases in the total nursing units of service for infants and children of 7 and 16 per cent, respectively.# This compares with an increase in nursing encounters in the Columbus County Children's Clinic of 34 per cent during the same period. Even if all WIC certification and recertification encounters with nurses are eliminated, Children's Clinic nursing activity would still show an increase of 25 per cent.

A number of fortuitous circumstances facilitated the Health Department's response to this increased demand. Following the opening of a new county hospital, the Health Department moved into the old hospital building. The availability of new space permitted the Children's Clinic to increase its operations from two to five days per week. Not only were the two full-time child health nurses able to spend more of their time in the Clinic, but a part-time nurse was hired to help out. In recognition of the contribution of the Health Department to child health in Columbus County, the County Commissioners increased local contribution to the Maternal and Child Health budget 20 per cent between FY78 and FY79. However, no special recruitment program was undertaken to increase WIC enrollment. The principal means for informing the public remained word of mouth.[7]

Since January 1979, a change in WIC reporting requirements has made it slightly easier for health departments to request funds for nursing staff equivalents. However, this can alleviate only part of the problem. WIC funds can only be used for reimburse for WIC-related services. The increasing percentage of children on WIC receiving non-WIC services from public health nurses suggests that these same children may be contributing disproportionately to the increase in all nursing encounters. If this is the case, consideration must be given to providing the additional resources necessary to handle this load. New mechanisms for supporting traditional child health activities in health departments taking on a WIC program may be indicated.

Acknowledgments

This report would have been impossible without the cooperation of the Columbus County Health Department, and especially of John R. Black, M.D., Health Director. The Children's Clinic staff who provided all the services and data were Ruth Storms, P.N.P., Marjorie Council, R.N., Hildred Scarborough and Ruth Long. WIC case-load data were provided by Judith Long and Sandra Roholt. Finally, none of the data would have been interpretable without the technical assistance and consultation of Sharon Schramm.

References

1. Foreman C. Hearings before the Subcommittee on Elementary, Secondary, and Vocational Education. U.S. Congress, House, Committee on Education and Labor, 92nd Congress, 2nd session, April 27, 1978; 184.

2. Edozien JC, Switzer, BR, Bryan RB. Medical Evaluation of the Special Supplemental Food Program for Women, Infants, and Children. Chapel Hill, N.C.: University of North Carolina, School of Public Health, Department of Nutrition, 1976.

3. Weiler PG, Stalker HP, Jennings SW, et al. Anemia as a criterion for evaluation of a special supplemental food program for women, infants, and children. Pediatrics 1979;63:584.

4. Food and Nutrition Committee. Report to the Urban Affairs Council, 1969. Cited in To Save the Children: Nutritional Intervention through Supplemental Feeding. U.S. Congress, Senate, Select Committee on Nutrition and Human Needs, 93rd Congress, 2nd session. Washington, D.C.: U.S. Government Printing Office. 1974:3.

5. Humphrey HH. Congressional Record. U.S. Congress, Senate, 92nd Congress, 2nd session, August 16, 1972;S28588.

6. U.S. Congress, Senate, Committee on Agriculture, Nutrition, and Forestry. The Child Nutrition Amendments: Hearings before the Subcommittee on Nutrition, on S-3805. 95th Congress, 2nd session, April 6, 1978;187–240.

7. Black JR. Personal commuinication, May 4, 1981.

# Maternal and Child Health Quarterly Reports, Maternal and Child Health Branch, North Carolina Division of Health Services.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Closure Among U.S. Community Hospitals, 1976–1980

A Description and a Predictive Model

ROSS M. MULLNER, PH.D.,[*] CALVIN S. BYRE, PH.D.,[*]
PAUL S. LEVY, SC.D.,[†] AND JOSEPH D. KUBAL, M.A.[*]

Closure of U.S. community hospitals during the years 1976–1980 is analyzed by describing the distributional patterns of closings among hospitals classified by certain institutional characteristics: bed size, number of facilities and services, ownership, teaching status, location in a Standard Metropolitan Statistical Area (SMSA) or in a non-SMSA, and location in one of the nine United States Census Divisions. The method of Cox regression analysis is used to show that each of these institutional characteristics is independently, and significantly, associated with closure, and to estimate the strength of the association. On the basis of this analysis, a theoretic model is constructed for estimating the relative probability of closure for any community hospital with a given set of these institutional characteristics.

THERE IS A GROWING AWARENESS among researchers that hospital closure in the United States is closely and significantly related to events and processes occurring in the health care industry in general, and that an understanding of hospital closure is requisite for fully understanding and dealing with such larger issues as the problem of reimbursement, the capability of health care institutions to adapt to changes in the communities they serve, excess hospital bed capacity, equal access to good quality health care and the distribution of physicians. For this reason, hospital closure has in recent years come to figure prominently in an increasing number of studies.[1] Intrinsic studies of closure are comparatively rare, however, apart from several studies that examine a few particular cases; these cases are instructive as examples but furnish evidence that is difficult to generalize. Yet it is clear that a solid body of descriptive information about closure, showing where and when it most frequently occurs, and a theoretic model capable of predicting its occurrence, are necessary before the implications of closure for other health care issues can be adequately explored and before effective policies can be devised for dealing with it.

One descriptive study on a fairly large scale was that of Ruchlin, Pointer and Cannedy, who in 1972 and again in 1973 examined the characteristics of nonfederal short-term hospitals registered with the American Hospital Association (AHA) that closed in the United States in 1969.[2–3] They compared the distributions of closed hospitals with those of hospitals that remained open within categorizations of hospitals by location in U.S. Census Divisions, by types of facilities and services offered, by bed size, and by control and

---

[*] Hospital Data Center, American Hospital Association, Chicago, Illinois.

[†] School of Public Health, University of Illinois at the Medical Center, Chicago, Illinois.

Address for reprints: Ross Mullner, Hospital Data Center, American Hospital Association, 840 N. Lake Shore Drive, Chicago, IL 60611.

0025-7079/82/0700/0699/$01.05 © J. B. Lippincott Co.

ownership (i.e., for-profit, state and local government, and voluntary). They also calculated "the mean value of select operating characteristics of failed and nonfailed hospitals" in order to isolate the variables that seemed most indicative of "operational viability." However, while their methologic approach was fresh and stimulating, their findings and conclusions were somewhat questionable and of limited usefulness. Of the 55 hospital closings that occurred in the single year that they considered, they were only able to deal with the 23 for which data were available. Moreover, they did not attempt to discover the relative strength of the association of the variables they examined with closure; hence, their model for predicting institutional viability is not capable of determining the relative probability that hospitals with given values for these variables will close.

McNeil and Williams, in "Wide Range of Causes Found for Hospital Closures,"[4] dealt with 231 U.S. hospitals registered with the AHA that closed between January 1975 and December 1977. The descriptive categories they used were similar to those used by Ruchlin et al. The focus of their article, however, was the causes of hospital closure in the U.S. In investigating this, their approach was not to construct an objective model of closure based on salient features of closed hospitals, but rather to classify and interpret the reasons given by hospital personnel themselves for the closings of their hospitals.

Mullner, Byre and Kubal, in "Closed U.S. Community Hospitals, 1972–1978: Perspectives and Trends,"[5] examined the 315 U.S. community hospitals registered with the AHA that closed within a seven-year period. This study was primarily descriptive, analyzing the distributions of closed hospitals within categories similar to those used by Ruchlin et al. and by McNeil and Williams, and noting how they vary from the distributions of all hospitals within the same categories.

It is the purpose of the present study to continue the lines of inquiry pursued in the three articles just mentioned. The present study employs similar categories, based on certain institutional characteristics, to describe the distributional patterns of closings among the AHA-registered community hospitals that closed in the U.S. within the five-year period from 1976 to 1980. It focuses on them because they form a large and relatively homogeneous class of hospitals (they included 5,851, or 83.4 per cent, of the 7,015 U.S. hospitals existing in 1978, the middle year of the period). Community hospitals are defined by the AHA as "all nonfederal short-term general and other special hospitals, excluding hospital units of institutions, whose facilities and services are available to the public."[6] Furthermore, this study uses multiple regression analysis to measure the strength of the association of closure with each of the institutional characteristics considered and, on the basis of this analysis, constructs a theoretic model for estimating the relative probability of closure for any community hospital with a given set of these characteristics.

The set of institutional characteristics by which hospitals might be classified and whose association with closure might be analyzed is vast. A subset whose members are well defined, determinate, objectively quantifiable and relatively stable has been selected for examination; previous research indicates that these characteristics have an important relation to closure. They are bed size, number of facilities and services, control (investor owned; nongovernment, not-for-profit; and government, not-for-profit), teaching status, location in a Standard Metropolitan Statistical Area (SMSA) or in a non-SMSA, and location in one of the nine United States Census Divisions.

This article is divided into two main parts. The first part begins with a brief descriptive overview of closure among all U.S. hospitals during the years from 1976 to 1980, dividing them into the two classes of community hospitals and noncommunity hospitals. The latter is a catchall, heterogeneous class comprising federal hospitals (psychiatric and general and other special) and nonfederal hospitals (including psychiatric hospitals, tuberculosis and other respiratory disease hospitals, and long-term general and other special hospitals). The rest of this part is devoted to closure among community hospitals. The second part performs the multiple regression analysis and constructs the predictive model.

All data used in this study are drawn from the AHA Membership, Publication, and Circulation File.

Description

Overview of Hospital Closure in the U.S., 1976–1980

Table 1 shows the number of U.S. hospitals registered with the American Hospital Association that closed and the number of beds lost as a result, by year and by classification as community or noncommunity. In general, closures and beds lost to closures were most numerous toward the beginning of the period, after which they declined, only to increase again at the end. A total of 296 hospitals closed together with their 29,491 beds. Although most of these hospitals were community hospitals (226), the bed losses were fairly evenly divided between the two classes, indicating that closed noncommunity hospitals were on the average much larger than closed community hospitals.

Community Hospital Closures, 1976–1980

Closings by Bed-Size Category. Table 2 shows for each bed-size category of community hospitals the number of hospitals and beds lost to closings during the five years and the proportion of all closed hospitals and closed beds that they represent. Calculated for each category is the rate of closure and of beds lost thereby per 100 hospitals and 100 beds in existence in 1978, the middle year of the period. For purposes of comparison, the number of community hospitals and beds existing in each category in 1978 is also shown, with the proportion of all community hospitals and beds that each represented.

It is apparent from Table 2 that, the smaller the bed-size category, the more subject its members are to closure: the rates of both closings and of their resulting bed losses decrease progressively with increases in bed size, dropping precipitously between the 100–199 and the 200–299 bed-size categories. Closure is, in fact, overwhelmingly a phenomenon of hospitals with fewer than 199 beds; making up 71.5 per cent of the nation's community hospitals and containing only 35.3 per cent of their beds in 1978, they accounted for 95.1 per cent of all community hospitals that closed during the five years and for 80.0 per cent of their bed losses due to closure.

Closings by Control Category. Table 3 shows the number, proportions, and rates of closings and resulting bed losses in each of the three control categories of community hospitals. The number of community hospitals and beds existing in each category in 1978 is shown, as well as the proportion of all community hospitals and beds that each represented.

Investor-owned (for-profit) community hospitals constitute the smallest category, with 12.5 per cent of all community hospitals and 8.3 per cent of their beds in 1978, but they accounted for the largest number of both closings (110, or 48.7 per cent) and of beds lost (6,876, or 42.2 per cent). Nongovernment, not-for-profit community hospitals, on the other hand, which make up the largest category, with 57.1 per cent of all community hospitals and 70.1 per cent of their beds in 1978, accounted for only 34.9 per cent (79) of closings and for only 38.3 per cent (6,228) of beds lost. State and local government community hospitals, the second largest group, with 30.4 per cent of hospitals and 21.6 per cent of beds in 1978, also had fewer closures (37, or 16.4 per cent) and somewhat fewer beds lost (3,171, or 19.5 per cent) than might be expected. The rates show similar relation-

TABLE 1. Closed Hospitals and Resulting Bed Losses by Hospital Type and Year of Closure

Hospital Type	Number of Closed Hospitals and Resulting Bed Losses											
	Year of Closure											
	1976		1977		1978		1979		1980		Total	
	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds	Number of Hospitals	Number of Beds
Community	52	3926	62	3943	37	2464	33	2721	42	3221	226	16,275
Noncommunity	18	4431	13	1660	16	2766	9	1687	14	2672	70	13,216
Total	70	8357	75	5603	53	5230	42	4408	56	5893	296	29,491

TABLE 2. Closed Community Hospitals and Resulting Bed Losses by Bed-Size Group

Bed Size	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978					
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Per Cent of Total Beds	Rate of Hospital Closure per 100	Rate of Bed Closure per 100
6– 24	30	13.3	511	3.1	268	4.7	5,077	0.5	11.2	10.1
25– 49	83	36.8	2,833	17.4	1,075	18.3	39,136	4.0	7.7	7.2
50– 99	62	27.4	4,218	25.9	1,444	24.7	103,460	10.6	4.3	4.1
100–199	40	17.7	5,463	33.6	1,398	23.9	196,652	20.2	2.9	2.8
200–299	7	3.1	1,546	9.5	717	12.2	173,751	17.8	0.9	0.9
300–399	2	0.9	698	4.3	392	6.7	132,828	13.6	0.5	0.5
400–499	1	0.4	432	2.7	240	4.1	106,352	10.9	0.4	0.4
500+	1	0.4	574	3.5	317	5.4	218,147	22.4	0.3	0.3
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7

ships and point to the same conclusion: for-profit community hospitals are by far the most subject to closure of the control categories, and are followed in the far distance by state and local government and by nongovernment not-for-profit hospitals.

Closings by Hospital Location. Table 4 shows the numbers, proportions, and rates of closings and resulting bed losses among community hospitals classified first according to location in the nine U.S. Census Divisions and then by location in SMSAs or non-SMSAs. The SMSA is defined by the U.S. Office of Management and Budget as a "county or group of contiguous counties that contain one or more cities of at least 50,000 population or twin cities with a combined population of 50,000 or more."

The rates of closures and bed losses reveal how disproportionately closings were distributed among the Census Divisions. The Pacific region, with rates of 7.0 for hospitals and 3.5 for beds, was the region whose community hospitals were most subject to closure. Having 12.5 per cent of all U.S. community hospitals and 10.9 per cent of their beds in 1978, it accounted for 23.0 per cent of their closings and 23.2 per cent of their bed losses through closure during the five-year period. The mid-Atlantic region also is noteworthy for having high rates and for disproportionately large numbers of closures and resulting bed losses. It contained 10.9 per cent of all community hospitals and 17.2 per cent of all community hospital beds in 1978, but accounted for 16.7 and 28.9 per cent of all community closures and beds lost, respectively. The West South Central region is interesting precisely because it is comparatively unremarkable: it had a slightly larger number of community hospital closures (37, or 16.4 per cent of the total) and a slightly lower number of beds lost to closures (1,421, or 8.7 per cent of the total) than might have been expected, given that it had 14.6 per cent of all community hospitals and 10.3 per cent of community hospi-

TABLE 3. Closed Community Hospitals and Resulting Bed Losses by Hospital Control

Hospital Control	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978					
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Per Cent of Total Beds	Rate of Hospital Closure per 100	Rate of Bed Closure per 100
Nongovernment—not-for-profit	79	34.9	6,228	38.3	3,339	57.1	683,258	70.1	2.4	0.9
State and local government	37	16.4	3,171	19.5	1,780	30.4	211,099	21.6	2.1	1.5
Investor-owned (for-profit)	110	48.7	6,876	42.2	732	12.5	81,046	8.3	15.0	8.5
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7

TABLE 4. Closed Community Hospitals and Resulting Bed Losses by U.S. Census Region and SMSA/Non-SMSA Location

Census Region SMSA/Non-SMSA Location	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978				Rate of Hospital Closure per 100	Rate of Bed Closure per 100
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Per Cent of Total Beds		
New England	7	3.1	468	2.9	259	4.4	51,015	5.2	2.7	0.9
Mid-Atlantic	37	16.4	4,736	29.1	636	10.9	167,869	17.2	5.8	2.8
South Atlantic	28	12.4	1,877	11.5	809	13.8	149,590	15.3	3.5	1.3
East North Central	23	10.2	1,809	11.1	910	15.6	193,224	19.8	2.5	0.9
East South Central	16	7.0	765	4.7	484	8.3	68,019	7.0	3.3	1.1
West North Central	24	10.6	1,369	8.4	799	13.7	98,972	10.2	3.0	1.4
West South Central	37	16.4	1,421	8.7	853	14.6	100,785	10.3	4.3	1.4
Mountain	2	0.9	62	0.4	362	6.2	39,424	4.1	0.6	0.2
Pacific	52	23.0	3,768	23.2	739	12.5	106,505	10.9	7.0	3.5
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7
SMSA	148	65.5	12,620	77.5	2,953	50.5	725,444	74.4	5.0	1.7
Non-SMSA	78	34.5	3,655	22.5	2,898	49.5	249,959	25.6	2.7	1.5
Total	226	100.0	16,275	100.0	5,851	100.0	975,403	100.0	3.9	1.7

tal beds in 1978. This is the region that, in the late 1960s and in the early and mid-1970s, accounted for the greatest number and the highest proportion of closed community hospitals.[3,4,5]

All other regions had fewer closings and resulting bed losses than might have been expected from the proportion of U.S. community hospitals and beds that they contained. The Mountain region had not only the fewest closings, but also the lowest proportion of hospitals and beds lost to closings.

As the rates of bed losses due to closings indicate, metropolitan and nonmetropolitan areas did not differ significantly in the proportions of their community hospital beds lost through hospital closings. There was, however, a significant difference in the proportions of their hospitals that closed: metropolitan hospitals, which made up half (50.5 per cent) of all community hospitals in 1978, accounted for 62.1 per cent of the closings.

Closings by Number of Facilities and Services. Table 5 shows the numbers, proportions, and rates of closings and resulting bed losses among community hospitals classified by numbers of facilities and services. As the columns with the rates show, by far the highest proportions of closings and of bed losses occurred among hospitals with only one to four facilities, and the proportions became lower as bed size increased.

Closings by Teaching/Nonteaching Status. Table 5 also shows the distributions of closings and of resulting bed losses among community hospitals classified according to their status as teaching or nonteaching hospitals. The AHA considers a hospital to be a teaching hospital if it meets any one of the following criteria: 1) it has a medical residency program approved by the American Medical Association; 2) it is a member of the Council of Teaching Hospitals; or 3) it is affiliated with a medical school. As can be seen from the table, teaching hospitals are much less subject to closure than are nonteaching hospitals. While the overall U.S. rate for hospital closure was 3.9, and for bed losses due to closure 1.7, the rates were only 1.1 and 0.6, respectively, for teaching hospitals, but 4.6 and 3.2 for nonteaching hospitals.

Discussion

In the preceding description of the distributions of closure in the various categorizations of community hospitals, it was assumed, as it was assumed by earlier researchers, that each of the institutional characteristics on which the categorizations are based bears a significant and unique associative relationship to closure. It was assumed, that is, that each characteristic affords a perspective onto a different and separate facet of the phenomenon.

It will be evident upon reflection, however, that this assumption is not necessarily correct. Through examination of the relationship between closure and each characteristic taken singly, the possible relationships that may obtain between the various characteristics themselves have not been taken into account. To take an example, among hospitals classified by control, the highest rates of closure were found among the investor owned; among hospitals classified by size, among the smaller bed-size groups. But investor-owned hospitals tend on the average to be smaller than those of the other control types; the high rate of closure among them may be due, therefore, not to some property that they possess by virtue of their form of control, but to some property that they possess by virtue of their size.

The validity of the assumption must, therefore, be tested. In order to discover whether, and how strongly, each of the institutional characteristics that has been considered is independently associated with closure, a multiple regression analysis was performed to obtain estimates of the relationship between closure and each

TABLE 5. Closed Community Hospitals and Resulting Bed Losses by Number of Facilities and Services and Teaching Status

Number of Facilities and Services, Teaching Status	Number of Closed Community Hospitals and Resulting Bed Losses (1976–1980)				Number of Total U.S. Community Hospitals and Beds Existing in 1978					
	Number of Hospitals	Per Cent of All Closed Hospitals	Number of Hospital Beds	Per Cent of Beds Lost to Closings	Number of Hospitals	Per Cent of Total Hospitals	Number of Hospital Beds	Percent of Total Beds	Rate of Hospital Closure per 100	Rate of Bed Closure per 100
—*	3	1.3	58	0.4	586	10.0	52,889	5.4	0.5	0.1
1– 4	69	30.5	2,033	12.4	312	5.4	10,221	1.1	22.1	19.9
5– 9	68	30.1	3,930	24.0	1,187	20.3	65,748	6.7	5.7	6.0
10–14	58	25.8	6,000	36.6	1,210	20.7	119,712	12.3	4.8	5.0
15–19	17	7.5	2,090	12.7	920	15.7	146,508	15.0	1.8	1.4
20–24	8	3.5	1,127	6.9	633	10.8	156,318	16.0	1.3	0.7
25+	3	1.3	1,155	7.0	1,003	17.1	424,007	43.5	0.3	0.3
Total	226	100.0	16,393	100.0	5,851	100.0	975,403	100.0	3.9	1.7
Teaching	14	6.2	3,464	21.1	1,278	21.8	567,660	58.2	1.1	0.6
Nonteaching	212	93.8	12,929	78.9	4,573	78.2	407,743	41.8	4.6	3.2
Total	226	100.0	16,393	100.0	5,851	100.0	975,403	100.0	3.9	1.7

* Hospitals shown on this line were nonresponders to the AHA Annual Survey of Hospitals; data on their facilities and services are not available.

characteristic, while controlling for confounding by the other characteristics. The data source used for the analysis was the 1976 AHA Annual Survey Tape; it lists all U.S. hospitals in operation as of September 30, 1975, the time immediately preceding the period covered by this study and at which all closed hospitals examined here were still open.

From this source and from the information on closed hospitals contained in the AHA Membership, Publication, and Circulation File, a working tape was constructed containing the following variables for each of the 5,859 hospitals in operation as of September 30, 1975:

Number of beds;
Whether or not it was a teaching hospital;
Whether or not it was located in an SMSA;
Number of facilities and services (0–52);
Whether or not it was investor owned;
Whether or not it was controlled by a state or local government;
Location by U.S. Census Division; and
Whether or not it closed between Jan. 1, 1976 and Dec. 31, 1980 and, if it closed, the month.

The procedure followed in the multiple regression analysis was developed by D. R. Cox and is commonly known as "Cox regression" or "life table regression."[7,8] It is now widely used by epidemiologists and biometricians for analyzing survival time as a function of a set of independent variables when the maximum time in which any individual can be studied is determined by the length of the study period. The method is appropriate in the present instance, in which the "individual" is the hospital and the study period is five years. The model specified by the procedure assumes that the chance of a hospital's closing during a particular period is a product of two functions, one depending only on time and the other depending only on the characteristics of the hospital. The model is specified as follows:

$$h(t) = \lambda(t)\, f(X_1, \ldots X_k), \qquad (1)$$

where $h(t)$ = the chance of a hospital's closing in t years, $\lambda(t)$ = the function of time, and $f(X_1, \ldots X_k)$ = the function of the hospital characteristics X_1, X_k.

The function $\lambda(t)$ is common to all hospitals, regardless of their characteristics. It is the function $f(X_1, \ldots X_k)$ that determines the effect of hospital characteristics on the risk of closure and that is, therefore, the regression equation to be estimated. The form of this function most often used in the Cox regression is the exponential given by:

$$f(X_1, \ldots X_k) = \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k), \qquad (2)$$

where $\beta_1, .. \beta_k$ are the regression coefficients to be estimated. They are estimated by the statistical procedure of *conditional likelihood* developed by Cox; this makes it possible to estimate the regression coefficients without explicitly formulating the function $\lambda(t)$.

The variables were specified as shown in Table 6. Number of beds and number of facilities and services are the only quantitative variables; the others are categoric. Location in U.S. Census Divisions has been coded as two dichotomous variables (the usual procedure in multiple regression analysis for categoric variables having more than two categories). The remaining variables are already dichotomous.

The variables were fitted to the exponential equation (2) by means of the Cox regression computer program known as PHGLM, which has recently been incorporated into the Statistical Analysis System (SAS) and is described in the SAS Supplemental Library Users Guide.[9]

Table 7 shows the magnitude of the estimated Cox regression coefficient for each variable and the statistical significance of the relationship between the variable and closure when there is control for possible confounding due to the other variables. As can be seen, each of the characteristics examined in the descriptive analysis does in fact have a significant independent association with closure. Thus, the Cox regression analysis validates the descriptive categories employed in this study.

TABLE 6. Variables

Variable	Coefficient	Specification
Number of beds (X_1)	β_1	Actual number of beds
Teaching status (X_2)	β_2	1 if teaching hospital; 0 if nonteaching hospital
Number of facilities and services (X_3)	β_3	Actual number of facilities and services
Investor owned (X_4)	β_4	1 if investor owned; 0 otherwise
Government controlled (X_5)	β_5	1 if controlled by a state or local government; 0 otherwise
Mid-Atlantic or Pacific location (X_6)	β_6	1 if located in Mid-Atlantic or Pacific Census Division; 0 otherwise
Mountain location (X_7)	β_7	1 if located in Mountain Census Division; 0 otherwise
SMSA location (X_8)	β_8	1 if located in an SMSA; 0 otherwise

The Theoretic Model

Not only does the Cox regression analysis validate the use of these institutional characteristics to form descriptive categories for investigating hospital closure, but, as will be seen, the technique can be expanded to create a theoretic model for estimating hospitals' relative risks of closing. First, odds ratios were calculated that estimate the risk of closure for hospitals having a given level of each variable (e.g., location in an SMSA) relative to hospitals having some reference level of the same variable (e.g., location in a non-SMSA) when the values of the other variables are the same. They were derived from the model as follows:

$$R = e^{\beta(X_1 - X_0)} \tag{3}$$

TABLE 7. Estimated Regression Coefficients and Relative Risks as Derived from Cox Regression Model

Variable	Specification	β Coefficient	Significance (p value)	Relative Risk Odds-Ratio Derived from Model
Beds	100 beds/300 beds	−0.0075	<0.0001	4.51
Teaching status	Teaching/nonteaching	1.0332	0.0008	2.81
Investor owned	Yes/no	1.0278	<0.0001	2.79*
SMSA	Yes/no	0.7993	0.0000	2.22
Coastal location	Mid-Atlantic and Pacific/other region	0.6389	0.0000	1.89†
Facilities	4 facilities/10 facilities	−0.0769	<0.0001	1.58
State or local government control	Yes/no	−0.4828	0.0231	0.62*
Mountain location	Mountain/other region	−1.5787	0.0272	0.21†

* Relative to nongovernment not-for-profit.
† Relative to nonmountain and noncoastal.

where R = relative risk of closure, X_1 = specified level of a variable X, X_0 = some reference level of a variable X, and β = regression coefficient for X as obtained from Cox regression. The relative risks estimated in this way control for confounding by the other variables being investigated.

The relative risk odds ratios shown in Table 7 indicate that, for example, a hospital having 100 beds has a 4.5 times greater risk of closure than does a hospital having 300 beds, *ceteris paribus*, an investor-owned hospital, 2.79 times greater than a nongovernment not-for-profit hospital, etc. Relative risks lower than 1.00 indicate a protective effect. For example, a hospital that is not government controlled has a 1.6 times greater risk of closure than does a government-controlled hospital (i.e., 1/0.62 = 1.61). Here it may be noted that, while the significance of the association noted above between teaching status and closure was expected, the *nature* of that association is somewhat surprising. It was shown in the descriptive analysis—in which there was no control for confounding effects—that teaching hospitals are less likely to close than are nonteaching hospitals. It is seen, however, that teaching status per se is associated with a higher likelihood of closure than is nonteaching status. This finding of an unexpected positive association should be regarded with caution, however, since six of the 14 teaching hospitals that closed were located in the New York City SMSA. That the association between teaching status and closure is positive rather than negative may be a fact that is peculiar to the present data set, due to unusual circumstances affecting the New York hospital industry during the period; it might not be the case in other years.

Next, the regression coefficients estimated as described above were used to calculate a score, S, for each of the 5,859 community hospitals considered in the study, based on the particular values of its characteristics, by means of the formula:

$$S = \beta_1 X_1 + \ldots + \beta_k X_k. \tag{4}$$

The correlations between the scores, S, of the hospitals and their rates of closure were then examined by grouping the hospitals into deciles on the basis of their scores and computing the five-year closure rates for each decile. As Table 8 shows, the scores generally increase with increases in the closure rates; this correlation is especially evident in the deciles with the higher scores and is most striking in the

TABLE 8. Prognostic Scores of Closed Hospitals 1976–1980 Compared With Number of Closings and Closure Rates

Decile*	Prognostic Scores(S)	Number of Closures	Five-year Closure Rate (per 100 Hospitals)
1st	−17.05 to −3.69	1	0.17
2nd	−3.68 to −2.73	3	0.51
3rd	−2.72 to −2.19	3	0.51
4th	−2.18 to −1.75	4	0.68
5th	−1.74 to −1.42	9	1.54
6th	−1.41 to −1.15	5	0.85
7th	−1.14 to −0.84	19	3.24
8th	−0.83 to −0.48	28	4.78
9th	−0.47 to 0.22	38	6.47
10th	0.23 to 2.37	116	19.83

* Five hundred eighty-six hospitals in all except the 10th decile, which contains 585 hospitals.

tenth decile, in which a five-year closure rate of 19.83 per cent is found among hospitals with scores of 0.23 to 2.37.[‡]

Thus, the score obtained for any community hospital can be interpreted as a prognostic index of its relative likelihood of closure.

Conclusion

This study has described the distributions of closed community hospitals within categories of hospitals classified according to the institutional characteristics of bed size, control, Census Division location, SMSA/non-SMSA location, number of facilities and services, and teaching status. It has demonstrated the validity of the assumption that these characteristics are intrinsically related to closure by means of the technique of Cox regression analysis. It has constructed from this method of analysis a theoretic model for estimating the relative probability of closure for any hospital having a given set of the characteristics.

It was not the purpose of this study to determine what the necessary or sufficient conditions of closure are, nor to discover its causes. It can be assumed that the closure of a hospital is typically the outcome of a long, dynamic process involving a number of complex, interrelated factors in the internal system of the institution (e.g., an inadequate number of medical staff, financial distress, underutilization, etc.) and in its external operating environment (e.g., competition from other institutions, changes in the population of the community served, etc.). The isolation of these and the discovery of the concatenations of causes and effects that are most often operative in hospital closure remain an important task for health care researchers to accomplish. Prerequisite for its accomplishment are, first, a comprehensive body of general information about the scope and dimensions of the phenomenon and, second, a series of prospective case studies examining the phenomenon *in process* in a large number of hospitals over a long period of time. It is hoped that this descriptive analysis has added to the former; the authors believe that this theoretic model, which makes it possible to identify those hospitals that are most likely ultimately to fall victim to closure, will facilitate the latter.

Acknowledgments

The authors would like to express their gratitude to Dale Matthews, Project Director, Hospital Data Center, American Hospital Association, and to Norman Iverson, Programmer Analyst, University of Illinois School of Public Health at the Medical Center, Chicago, for their expert assistance in writing the computer programs needed for this study.

References

1. Institute for Health Planning. Capacity reduction through facility closure and conversion: a bibliography with abstracts. Prepared for the Health Resources Administration, HRA 232-79-0036. Madison, Wis.: Institute for Health Planning, April 1980;6-18.

2. Ruchlin HS, Pointer DC, Cannedy LL. Hospitals that fail are often the small ones. Mod Hosp 1972;120:61.

3. Cannedy LL, Pointer DD, Ruchlin HS. Viability and hospital failure: methodological considerations and empirical evidence. Health Serv Res 1973;8:27.

4. McNeil D, Williams R. Wide range of causes found for hospital closure. Hospitals 1978;52:76.

5. Mullner R, Byre C, Kubal J. Closed U.S. community hospitals, 1972–1978: perspectives and trends. Soc Sci Med 1980;14D:355.

6. American Hospital Association. Hospital Statistics, 1980 Edition. Chicago: American Hospital Association, 1981;ix.

7. Cox DR. Regression models and life tables. J R Statist Soc Series B 1972;34:187.

8. Lee ET. Statistical methods for survival data analysis. Belmont, Cal.: Lifetime Learning Publications, 1980;306-18.

9. Harrell F. The PHGLM procedure. In: SAS Supplemental Library Users' Guide 1980 Edition. Cary, N.C.: SAS Institute, Inc, 1980;119-31.

[‡] Teaching status was included in the calculation of the scores shown, although, as mentioned above, the regression coefficient that was calculated for that variable might be valid only for the present data set. The scores were also calculated while leaving teaching status out of account, but they did not change appreciably.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Self-Care Education

Impact on HMO Costs

DONALD W. KEMPER, MSIE, M.P.H.*

The ability of a medical self-care program to have an impact on the utilization and cost of in-clinic and referral visits in a prepaid group practice was studied by following the medical encounter records of 900 HMO members in 217 families during a 12-month period. A control group selected by randomization after acceptance of an initial invitation to participate was employed. In spite of the program's impressive effect on both self-care knowledge and per-visit costs, the study could not demonstrate any significant program effect either on the frequency or on the total costs of clinic visits.

HEALTH EDUCATION is a mandated service for federally qualified Health Maintenance Organizations (HMOs). In setting that requirement, Congress contended that to maintain good health and optimal use of health care services, a person must be motivated and knowledgeable in making health-related decisions. Although most HMOs employ enough educational activity to meet minimum regulatory levels, few have embraced the spirit with which the requirement was imposed.

HMO managers have been reluctant to make substantial investments in educational strategies without a base of factual evidence indicating bottom-line cost reductions.

The impact of a medical self-care education program on the frequency and cost of health care services received by its participants in a prepaid group practice HMO is examined. Medical self-care education in this article refers to general educational efforts addressing the person's need to make decisions and apply home treatments regarding a broad range of common minor illnesses and injuries. The evaluation employs a control group selected by randomization after acceptance of an initial invitation to participate.

Background and Previous Findings

Although the overall self-care/self-help movement has many varied and early roots in the United States,[1,2] Keith Sehnert's *Course For the Activated Patient*[3] was the first of a new generation of medical self-care programs to address a broad range of health concerns and a basic change in the patient's role in the doctor/patient relationship. In the ten years since its 1970 introduction, similar classes and workshops have been organized in virtually all states of the nation, and self-care books, guides and videotapes have been developed to help meet the public's increasing interest in regaining self-responsibility for their health.[3–13]

Early evaluation of the Activated Patient program with both Medicaid and Medicare participants was hampered by methodologic and procedural problems

---

* Healthwise, Incorporated, Boise, Idaho.

Supported jointly by Grant R18 HS 02109 from the National Center for Health Services Research and by a grant from the W. K. Kellogg Foundation.

Reprint requests: Donald W. Kemper, Executive Director, Healthwise, Incorporated, PO Box 1989, Boise, ID 83701.

0025-7079/82/0700/0710/$00.95 © J. B. Lippincott Co.

that neither confirmed nor refuted its value.[4] The evaluation did suggest, however, that people would enroll and attend the classes and that their confidence about self-care increased.

Similarly, separate evaluations of 1975 self-care classes in rural Idaho and Oregon communities revealed substantial improvements in self-care knowledge and high enrollment and attendance levels.[14–15] Conclusions were positive and optimistic, but no direct measures of program effects on cost or utilization were obtained.

More recently, an evaluation of a self-care program for university students with upper respiratory infection symptoms suggested substantial savings in both visits and costs due to the program.[16] Other studies of categoric self-care programs involving asthma patients and home throat culture kits have contributed further to the evidence that self-care efforts can yield credible returns.

Two studies that evaluate the impact of the popular self-care book *Take Care of Yourself*[12] on physician use rates have been reported. Two physicians at the University of Washington compared patients' actual decisions to see a doctor with the actions recommended by algorithms in the book.[17] For the eight symptom categories studied, strict adherence to the algorithms would have increased the number of visits over that observed for five categories. Decreased utilization was predicted only for cases of hoarseness. As noted by the researchers themselves, generalizations based on the results are limited.

A second evaluation of *Take Care of Yourself* involved a prospective randomized control trial in a prepaid health plan.[18] The study followed the physician utilization of 1,300 families divided into three groups. Group 1 was the control group. Group 2 members received the book and were invited to a one-hour self-care seminar. Group 3 received the book, the seminar invitation and a $50 refund if the family's office visits dropped by at least one third during the six-month trial period The study found widespread use of the book within groups 2 and 3, and a larger decrease in clinic use rates for those groups than for the controls. However, statistical significance of the impact could not be demonstrated.

Although these previous attempts to evaluate self-care education in HMOs and other settings have generally returned positive and optimistic conclusions, they have fallen short of the direct financial evidence needed to justify large scale investment in general (as opposed to categoric) medical self-care initiatives. The present study was designed to help fill that need.

Description of the Program

The self-care program involved a series of ten two-hour workshop sessions led by a nurse practitioner. The format was informal, with emphasis placed on skill building and peer reinforcement through demonstrations and discussions. Active participation in the workshops was encouraged as the key method of developing the confidence needed to apply the skills and knowledge presented. Greater confidence in dealing with health concerns was viewed as equal to greater self-care competence in the overall purpose and goals of the program.

The program content (see Fig. 1) began with vital sign- and symptom-monitoring skills. Participants were taught to perform home physical examinations on themselves and their children. Subsequent sessions dealt with the management of abdominal problems, backaches and headaches, upper respiratory problems, skin problems and minor injuries. The topic selection was based on the frequency of visits reported in the National Ambulatory Medical Care Survey[20] and on participant interest, expressed in previous versions of the program. Discussions also emphasized

- The Home Physical Exam
- Abdominal Health Problems
- Backaches and Headaches
- The Common Cold
- Upper Respiratory Infections
- Skin Problems
- Injuries
- Emergencies
- Dental Care & Nutrition
- Life Style and Health

FIG. 1. Ten topics on family health.

a more activated relationship between the person and his/her health care provider.

Later in the series, the topics shifted away from minor illness and injury. Participants were told that although practical skills in those areas may save them money and improve their self-confidence in dealing with health problems, long-term improvements in health would depend more on broader life-style changes. Accordingly, the last few workshop sessions focused more on stress management, nutrition and fitness.

Each participant was given a 250-page self-care guide, *The Healthwise Handbook*,[8] and assigned readings prior to each session. The guide became a continuing reference source after the workshops were over. Consistency among workshop groups was maintained through the use of an instructor's guide[5] and a series of videotapes supporting the content and concepts of the program. Distribution of the support materials has allowed program replication in more than 15 states.

Evaluation Methods

Members of a small (6,000-person enrollment) prepaid group practice health plan in Boise, Idaho were contacted first by letter and then by phone to learn of their interest in attending a new health education class to be offered. Those who said yes (48 per cent) were randomly divided into experimental and control groups. Experimental group members were then encouraged to attend the actual workshop. Controls were asked to complete a questionnaire and told that although the trial workshops were filled, they would be recontacted if and when additional classes were offered. As shown in Figure 2, 59 individuals (55 per cent of the experimental group) actually participated in at least one workshop. Because of the program's emphasis on childhood and young adult problems, only those families with children under 18 years of age were asked to participate.

Although only 59 of the 107 families actively participated, the entire experimental group was used in the analysis to assure equivalency with the controls (Fig. 2). Demographic comparisons between the two groups revealed no apparent anomalies (Table 1).

The evaluation covered three general concerns: 1) the logistic success of the program; 2) the program impact on the predisposing factors of self-care knowledge and attitudes; and 3) the program impact on clinic utilization and cost.

Logistic success was measured by enrollment penetration rates, attendance and participant ratings of the workshop.

Self-care knowledge gains were measured by a 69-item self-care test designed around the specific learning objectives of each workshop session. The instrument was refined through several trial tests and professional reviews before being used. A split-half reliability test using the Spearman–Brown prophecy formula resulted in an R value of 0.94.[21]

Data on the utilization and cost of services by both the experimentals and controls were obtained from a computerized summary of clinic and referral encounters during a 12-month period. Cost data were based on the nonmember fees of the HMOs for the services received. The data

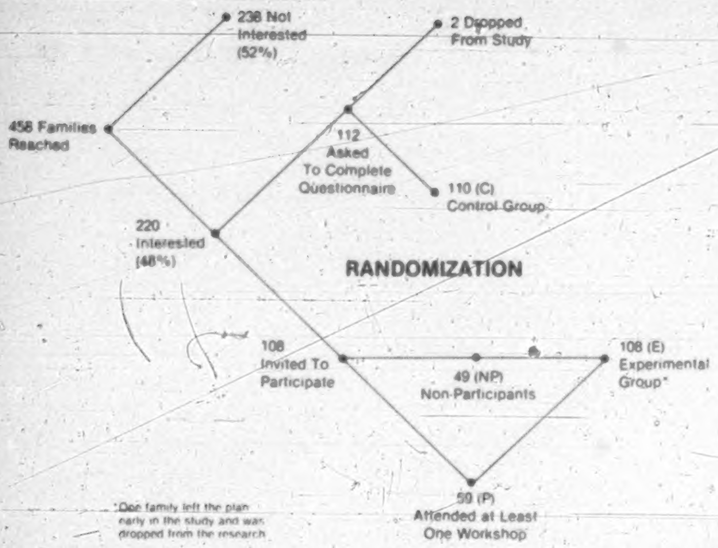

458 Families Reached
238 Not Interested (52%)
220 Interested (48%)
112 Asked To Complete Questionnaire
2 Dropped From Study
110 (C) Control Group
RANDOMIZATION
108 Invited To Participate
49 (NP) Non-Participants
108 (E) Experimental Group*
59 (P) Attended at Least One Workshop
*One family left the plan early in the study and was dropped from the research

FIG. 2. Experimental and control group selection.

were part of the routine financial management system of the HMO. Encounters were assigned to the three categories of preventive, minor or other, based on the symptom/problem codes. The International Classification of Health Problems in Primary Care (ICHPPC) system was used.[22] The "minor" category was designed to include most of the problem classifications that the self-care program directly addressed.

Analysis of the utilization and cost data was based on the statistic

$$X_i = \frac{\text{number of visits by family of participant } i \text{ during the 12-month period}}{\text{number of family members in family } i}$$

The t-test was used to test for significant differences.

Findings

Findings in the first two areas of evaluation were positive and highly consistent with results from previous studies discussed earlier. After adjusting for the control group, 26 per cent of families contacted had one or more members who participated in the workshop. Those who did participate attended an average of 5.5 of the ten sessions.

Eighty-three per cent of participants said the workshop improved health care for their families, and 81 per cent reported they had read at least half of the handbook.

Self-care knowledge scores (Table 2) increased 125 per cent for the participants and dropped only 1 per cent during the following six months. Scores for the control group increased only 8 per cent during the same period. The score improvement difference between the participants and controls was highly significant at the 0.01 level ($t = 9.53$).

Impact on Utilization

Findings regarding the program impact on utilization frequency of clinic services were less definitive. As indicated in Table 3, the experimental group averaged slightly more visits to the HMO clinic than did the controls. However, when referral visits are considered, the total frequency of visits is slightly higher for the controls. These differences are small and do not approach statistical significance. (The largest t value was for referral visits, $t = 1.26$.)

Additional analysis of the utilization data by calendar quarter, by demographic subgroups and for children only demonstrated neither statistical significance nor strong trends. A Kolmogrov–Smirnov test[23] did not refute the hypothesis that the two populations are identical.

Impact on Cost

The analysis of financial impact of the program included study of both costs per family member and cost per visit.

As indicated in Table 4, both the participating and nonparticipating elements of the experimental group accrued less cost over the study period than did the controls. Although the average cost differences reached as high as 24 per cent for referral and "other" clinic services, the high variances encountered prevented any statistically convincing findings. Only for the 42 per cent difference in referral cost averages between workshop participants and controls was statistical significance approached (t = 1.654).

Findings in the cost-per-visit analysis were more definitive. As shown in Table 5, the $2.54 per-visit difference between experimentals and controls was found to be significant at the 0.01 level (t = 2.67). When the per-visit costs were broken down into clinical, laboratory and x-ray components, only the $1.19 difference related to laboratory costs remained significant (t = 2.55). The apparent 28 per cent savings in laboratory cost and the 24 per cent savings in x-ray accounted for 73 per cent of the overall cost/visit margin.

As in the cost-per-visit-person analysis, it was found that both the experimental participants and the non-participants showed significantly lower costs/visits than did the controls (t = 2.06 and 2.16, respectively).

The overall cost per family of in-clinic and referral costs as shown in Table 6 reveals an average cost savings of $55.48 per family, or $5,975 for the experimental group. However, since only 59 of the 107 experimental group families participated in the workshop, the total average savings per participant was $101.27. It should be noted that the cost data applies to the 12-month period from December 1977 through November 1978.

TABLE 1. Comparison of Experimental and Control Groups, Prepaid Group Practice

	Experimental	Control
Number completing demographic questionnaire*	70	93
Age (mean)	35.4	36.1
Education (mean years of schooling completed)	14.5	14.8
Children (number under 18 years of age)	2.3	2.3
Sex		
Per cent female	86	72
Per cent male	14	28
Income		
Per cent < 5,000	0	1
5,000–9,999	8	11
10,000–14,999	37	33
15,000–19,999	19	26
Per cent > 19,999	29	24
Unknown	13	5
Pretest		
Knowledge (mean score)	28.3	27.3
Attitude (mean score)	7.7	7.0
Satisfaction (mean score)	7.0	6.8

* Questionnaire response rate averaged 75 per cent. Because utilization and cost data were available on all of the randomized subjects, the analysis in those areas was based on a larger sample.

Program Cost

Excluding program development and research expenditures, the program cost totaled approximately $65 per participating family. As shown in Table 7, 58 per cent of the cost was due to personnel and indirect costs, while direct expenditures for workshop supplies and handbooks amounted to $17 per family. Although some economies of scale would result from large-scale implementation of the program, the total cost would not be significantly changed.

TABLE 2. Self-Care Knowledge Scores

Self-Care Knowledge Scores	Experimental Group		Control Group
	Participants	Non-participants	
Pretest (n)	28.5 (59)	27.1 (11)	27.3 (93)
Posttest (n)	65.1 (40)	NA	NA
Per cent change*	124% ↑	NA	NA
Six-month test (n)	66.0 (34)	31.2 (9)	30.2 (71)
Per cent change from pretest†	111% ↑	8% ↑	8% ↑
Per cent change from posttest†	1% ↓	NA	NA

* Comparison of pretest to posttest scores of only those returning both questionnaires.

† Comparison of pretest to six-month and posttest to six-month scores of those returning all questionnaires.

Discussion

The primary question regarding self-care programs is whether, in the long run, their participants benefit, either by better health or by lower medical expenses. This study was unable to address the first half of the question directly because of resource limitations and the difficulty of assessing health-status changes. However, the cost issue was addressed, at least in part, and a number of secondary and logistic questions were answered with relevance to both cost and health-status concerns.

Logistically, it was demonstrated that large numbers of people from a general clinic population could be recruited and trained in medical self-care at a reasonable cost. This is a necessary but certainly not sufficient condition for program cost effectiveness. However, by enrolling more than 25 per cent of the HMO families contacted at a single point in time, the project reaffirmed the opportunity that self-care programs have for affecting large segments of the population.

Although health care knowledge and awareness are not directly related to health status, they are generally accepted as positive predisposing factors that health improvement programs should address. The study results clearly reinforce earlier findings that program participants do learn substantial amounts through self-care workshops and classes. In addition, these findings indicate that the retention of such knowledge can be high, at least for the six months following the program's completion.

TABLE 3. Visits/Family Member/Family/Year*

Visit Category	Experimental E			Control (C)	t (C–E)
	Participants	Nonparticipants	Combined (E)		
Number of families	59	48	107	110	
In-clinic visits					
Minor	2.68	2.05	2.40	2.22	−0.72
Preventive	0.87	0.63	0.76	0.75	−0.06
Total	4.31	3.67	4.03	3.99	−0.08
Referral visits	0.28	0.37	0.322	0.423	1.26
Total visits	4.59	4.04	4.35	4.41	

$$^{*}\ \frac{\text{Number of visits by family of participant } i \text{ during 12 months}}{\text{Number of family members in family } i}$$

TABLE 4. Cost/Family Member/Family/Year

Cost Area	Experimentals			Controls (C)	Significance Tests							
					E vs C		P vs C		NP vs C		P vs NP	
	Participants (P)	Nonparticipants (NP)	Combined (E)		Per Cent Change	t	Per Cent Change	t	Per Cent Change	t	Per Cent Change	t
In-clinic visits												
Minor	45.50	36.74	41.57	38.48	+8.0	—	+18.2	—	−4.5	—	+23.8	—
Preventive	19.21	12.53	16.21	17.19	−5.7	—	+11.7	—	−27.1	—	+53.3	—
Other	23.65	21.85	22.84	30.31	−24.6	—	−22.0	—	−27.9	—	+8.2	—
Total	88.35	71.11	80.62	85.98	−6.2	—	+2.8	—	−17.3	—	+24.2	—
Referrals	28.49	47.88	37.19	49.23	−24.5	1.015	−42.1	1.645*	−2.7	—	−40.5	1.13
Totals	116.85	118.99	117.81	135.21	−12.9	—	−13.6	—	−12.0	—	−1.8	—
Number of families	59	48	107	110								

* Significant at 0.1 level.
t values less than 1.00 omitted.

TABLE 5. Cost/Visit/Family by Cost Element

Cost Element	Experimentals			Controls (C)	t = Test Values			
	Participants (P)	Nonparticipants (NP)	Combined (E)		E vs C	P vs C	NP vs C	P vs NP
Clinical cost/visit	16.07	15.13	15.64	16.34	−1.17	−.313	−1.71	+1.05
Lab cost/visit	3.01	3.22	3.10	4.29	−2.55*	−2.20*	−1.67	−.41
X-ray cost/visit	1.87	2.39	2.10	2.77	−1.35	−1.48	−.60	−.76
Total cost/visit	20.96	20.74	20.86	23.40	−2.67*	−2.06*	−2.16*	+.18
N	59	48	107	110				

*Significant at 0.01 level.

The utilization data were both unexpected and enlightening. Based on the earlier findings by Moore and the program emphasis placed on reducing inappropriate utilization, a substantial reduction of in-clinic visits had been expected. It had been predicted that little impact would be seen in the specialty referral areas. In fact, the reverse appears more true. The observed reduction in referrals for the participants may be explained by one or more of the following speculations:

1. Better self-care applied earlier in the course of an illness led to fewer complexities of the illnesses encountered; thus, fewer specialty services were needed.
2. The program further developed the participants' confidence in their primary care professionals so that demands for referrals were reduced.
3. Statistical aberration—a larger sample would have refuted the findings (the finding was not statistically significant).

Although the favorable findings in the reduction of per capita ambulatory costs are not statistically convincing, the program impact on per-visit costs is. The definitive finding that self-care classes lead to lower laboratory costs per visit is, however, difficult to explain. Although workshop content includes encouragement of patient questioning of the purposes of laboratory tests and x-rays, large cost differences in those areas were not expected.

Also difficult to explain are the consistent differences between the experimental nonparticipants and the controls. Since neither group received any more "treatment" than a brief program description, the differences must be explained by a basic population bias. The classic randomization of the initial experimental and control groups would suggest that experimental group individuals declining to participate tended to use fewer and less expensive services than those who chose to come to the workshops. Such a self-selection bias would tend to strengthen the cost-effectiveness benefits of the program.

TABLE 6. Cost per Family

Cost Area	Experimental ($)	Control ($)	Savings ($)
In-clinic visits	333.05	353.33	20.28
Referrals	147.39	182.95	35.56
Total	480.44	536.28	55.84
Number of families	107	110	

Total savings = 107 × 55.84 = $5,974.88; savings/workshop participant: $5,974.88 ÷ 59 = $101.27.

Conclusions

It can be concluded from the findings that the self-care program caused a reduction in the cost per visit incurred by program participants. Further, it was shown that more that 25 per cent of the young families in a clinic population voluntarily participated in the program and that their knowledge of the appropriate management of common health problems increased significantly.

Although both clinic and referral costs were lower for the experimental group than for the controls, the high variance encountered prevents conclusions either de-

TABLE 7. Program Replication Cost

Instructors	
Three nurse practitioners (20 class hours each)	$ 875
59 Healthwise Handbooks, $11* each	649
59 sets of supplies, $6 each	354
3 playings of Healthwise videotapes, $1,000 × 30%	300
10 days of health educator's time to coordinate classes, $70	700
Recruitment costs (estimated)	300
Total direct costs	$ 3,178
Indirect support cost (20%)	636
Total cost	$ 3,814
Total cost per participant	$ 64.64

* Now available at quantity discount for $4.64.

nying or confirming a favorable impact on total costs.

Recommendations

The findings of this study point directly to the urgent need to develop improved self-care education programs and to evaluate more fully their effect on health status and economic benefits:

1. Future research into the impact of self-care programs on clinic utilization should attempt to reduce the effect of high variance on the analysis by consideration of family size and previous health history in the design of the study statistic. (Example: $x = b$ visits $+ b_1$ family size $+ b_2$ health status.)
2. Sample size should be adequate to demonstrate statistical significance if the sample means differ by 10 per cent or more.
3. Long-term evaluations covering three or more years should be attempted, to measure any delayed changes in knowledge, attitude and utilization.
4. Stable clinical settings should be selected to minimize both the number of out-of-clinic visits and the turnover of the medical staff.

References

1. Green LW, Werlin SH, Schauffler HH, Avery CH. Research and demonstration issues in self-care: measuring the decline of mediocentrism. Health Educ Mongr Summer 1977.

2. Levin LS, Katz AH, Holst E. Self-Care Lay Initiatives in Health. New York: Prodist, 1976.

3. Sehnert K, Nocerino JT. The Activated Patient: A Consumer-Oriented Program on Preventive Medicine and Self-Help Medicine. Washington, D.C.: The Center for Continuing Health Education, Georgetown University, 1977.

4. The Center for Continuing Health Education. A Course and Evaluation of The Activated Patient Concept. Washington, D.C.: The Center for Continuing Health Education, Georgetown University, 1977.

5. Healthwise, Incorporated. Healthwise Instructor's Guide. Boise: Healthwise, Incorporated, 1979.

6. Georgetown University. Medicaid Citizens as Activated Patients: A Course Guide. Washington, D.C.: Georgetown University, September 1976.

7. Reeves J. Health Skills for Parents. Health Skills Associates, 1975.

8. Roberts T, Tinker K, Kemper D. Healthwise Handbook. Boise, Idaho: Healthwise, Incorporated, 1979.

9. Samuels M, Bennett H. The Well Body Book. Clinton, Mass.: Colonial Press, Inc, 1973.

10. Sehnert K. How To Be Your Own Doctor (Sometimes). New York: Grosset & Dunlap, 1975.

11. Sehnert K. The Patient as a Paramedical. Virginia Medical Monthly 1972;99:409.

12. Vickery DM. Take Care of Yourself: A Consumer's Guide to Medical Care. Reading, Mass.: Addison–Wesley Publishing Company, 1976.

13. Vickery DM, Fries JF, Pantell RH. Taking Care of Your Child. Reading, Mass.: Addison–Wesley Publishing Company, 1977.

14. Grant RH. Family and self help education in isolated rural communities. Health Educ Monogr 1977;5:145.

15. Kemper, DW, Tinker K. Common Sense–Common Health Evaluation. Boise: Healthwise, Incorporated, 1975.

16. Apka J, Averill BW. Self care for colds: a cost-effective alternative to upper respiratory infection management. Am J Public Health 1979;69:814.

17. Berg AO, LoGerfo JP. Potential effect of self-care algorithms on the number of physician visits. N Engl J Med 1977;300:535.

18. Moore SH. Randomized Controlled Trial of a Self Care Book. MPH thesis, University of Washington, Seattle, Washington, 1978.

19. Green LW, Drenter MW, Deeds SG, Partridge KB. Health Education Planning. Palo Alto, Calif.: Mayfield Publishing, 1980;8.

20. National Center For Health Statistics. National Ambulatory Medical Care Survey, Preliminary Data From the 1973–1974 Survey, July 15, 1975.

21. Downie NM, Health RW. Basic Statistical Methods. New York: Harper & Row, 1974.

22. World Organization of National Colleges, Academics and Academic Associations of General Practitioners and Family Physicians. International Classification of Health Problems in Primary Care. Chicago: American Hospital Association, 1975.

23. Blalock HM. Social Statistics. New York: McGraw–Hill, 1960.

MEDICAL CARE
July 1982, Vol. XX, No. 7

Reimbursing for Residency Training

How Many Times?

MOHAN L. GARG, SC.D.,* MOUNIR ELKHATIB, M.D.,† WARREN M. KLEINBERG, M.D., M.P.H.,‡ AND JACK L. MULLIGAN, M.D.§

Through an investigation of charges generated for 105 patients treated for congestive heart failure, three issues of reimbursement policy for graduate medical education are examined: 1) reimbursement policy by third parties for resident salaries and for the portion of teaching physicians' time devoted to graduate medical education; 2) the level of involvement for physicians in dual roles as attending physicians for private patients when simultaneously engaged in graduate medical education and supervising patient care delivered by residents; and 3) the greater utilization of hospital resources in teaching units. The results demonstrate that the reimbursement policy, as presently practiced, ignores the basic differences in the level of involvement between teaching and nonteaching physicians. The data also show for the first time the incremental effect of graduate medical education in a teaching hospital. Several alternatives are suggested to eliminate the deficiencies.

IN RESPONSE TO FISCAL PRESSURES, policymakers are seeking areas in which to contain health care costs. The Health Care Finance Administration, the federal agency responsible for Medicare reimbursement, has started to look at reimbursement policies for physicians involved in teaching and patient care. However, precise data to determine the differences in level of involvement between teaching and private physicians have not been reported in the literature. Some data suggesting excessive utilization of ancillary services and increase in length of stay for patients under the care of residents have been reported.[1,2] The two studies have neither a homogeneous patient population for comparison nor a control on attending physicians.

A study was carried out to compare the components of cost of congestive heart disease patients treated by residents under supervision of teaching physicians in a teaching unit and those treated by attending physicians in a private unit of the same hospital. The study looked at two issues: 1) the greater number of diagnostic tests ordered by the medical resident; and 2) the level of involvement of teaching physicians in comparison with that of private physicians for managing the same type of patient. Both issues have implications for the effect of a graduate medical education

---

* Department of Medicine, Medical College of Ohio, Toledo, Ohio.

† Department of Family Medicine, Medical College of Ohio, Toledo, Ohio.

‡ Department of Pediatrics, Medical College of Ohio, Toledo, Ohio.

§ University of Missouri School of Medicine at Kansas City.

Supported by a grant from the National Fund for Medical Education, Hartford, Connecticut.

Reprint requests: Mohan L. Garg, Sc.D., Division of General Internal Medicine, Department of Medicine, Medical College of Ohio, C.S. # 10008, Toledo, OH 43699.

0025-7079/82/0700/0719/$00.90 © J. B. Lippincott Co.

program on hospital rate setting in teaching hospitals.

Under the present reimbursement policy for hospital care by third-party insurers (Medicare, Medicaid and Blue Cross), the salaries and fringe benefits of residents and fellows, and the time devoted by the teaching physician to supervise the graduate medical education are considered general operating expenses of the hospital. Stated another way, physician employee expenses tend to increase per unit cost of each of the services produced by the hospital. For example, the daily room rates for the intensive care unit, coronary care unit, and semiprivate and private rooms include a specific amount that is generated because of teaching programs. In addition, the relative charges of producing some ancillary services also include a teaching-generated component. In short, every unit of service cost produced in a teaching hospital contains a portion that is due to the existence of teaching programs. This has been considered to be equitable by the Graduate Medical Education National Advisory Committee (GMENAC)[3] and the Institute of Medicine (IOM)[4] because residents purportedly spend 80 per cent or more of their time in patient care activities.

Materials and Methods

The data for the present study were collected from a community hospital associated with a medical school. All patients with a diagnosis of congestive heart disease during the 12 months of 1978 were selected for the study. Patient bills were gathered from the business office, and all diagnostic tests and procedures, along with their charges, were recorded. Charges for pharmacy and other supply items were also recorded. The data were then correlated with the medical charts to determine the day tests were ordered and the source of each order, whether a resident, attending physician or standing order of the hospital. Patients were also identified by location, the teaching unit or the private unit. All of the data were compiled for each patient ($n = 105$), each patient being given a coded identification number to assure confidentiality. The data were then keypunched and stored in the computer for subsequent analysis.

Unlike previous studies, strict controls on patient comparability were maintained, and all patients had a primary diagnosis of congestive heart failure at discharge confirmed by the medical records. Other patient characteristics such as sex, age and mode of payment were gathered to examine the group for homogeneity. Severity was controlled for by the method of the New York Heart Association Criteria Committee.[5]

Results

Of 105 patients with congestive heart disease there were 52 men and 53 women. Eighty-three were 65 years of age and over, and 22 were under the age of 65. Forty patients were treated in a teaching unit, 65 in a private unit. All of the patients in the private unit had associated secondary conditions, and 95 per cent of those in the teaching unit had secondary conditions. There was no significant difference in severity of congestive heart failure or of the associated conditions between study groups, nor was there a difference in mode of payment between groups.

Table 1 presents a summary comparison of teaching versus private unit patients for various categories of hospital services use. On an average, charges for diagnostic tests are 52 per cent higher for teaching patients, $1,414 for patients in the teaching unit as compared with $932 for patients in the private unit. The differences vary widely among laboratory departments, microbiology being almost three times higher and respiratory therapy and cardioperfusion being two times higher for teaching unit patients. In Table 2, average

TABLE 1. Average Charges for Patients With Congestive Heart Disease in Teaching and Private Units

	All Patients	Type of Patient		Per Cent Increase or Decrease in Charges
		Private	Teaching	
Number of patients	105	65	40	
Average charge ($)				
Total	2969	2820	3210	+14
Room	1641	1704	1538	−10
Pharmacy	134	113	169	+50
Supplies	78	72	89	+24
Diagnostic tests	1116	932	1414†	+52
Hematology	50 (5.6)*	46 (5.3)	56 (6.1)	+22
Microbiology	40 (3.5)	23 (1.9)	68 (6.0)	+196
General chemistry	322 (20.5)	266 (15.9)	411 (28.0)	+55
Cardiovascular	129 (3.8)	136 (3.7)	118 (4.1)	−13
Radiology	123 (5.4)	108 (4.8)	148 (6.4)	+37
Respiratory therapy and cardio-perfusion	304 (10.5)	218 (8.0)	444 (14.4)	+104
Miscellaneous	148 (6.3)	135 (5.4)	166 (7.7)	+23

* Average number of tests.

† Figures may not total accurately due to rounding error.

charges per patient for teaching and private unit patients with the same attending physicians are shown. Even though the attending physicians are the same, diagnostic and treatment charges are 64 per cent higher in the teaching unit. The three most excessively utilized areas in a teaching unit are microbiology, cardiovascular, and respiratory therapy and cardio-perfusion.

Since all of the patients have comparably severe congestive heart failure, the explanations for the differences in treatment must be found in the source of the treatment orders. The data in Table 3 show the

TABLE 2. Average Charges per Patient With Congestive Heart Disease for Private and Teaching Units With Same Attending Physician

	Private Unit	Teaching Unit	Per Cent Increase (+) or Decrease (−) in Charges
Number of Patients	20	21	
Average total charges ($)	2912	3716	+28
Room	1673	1714	+2
Pharmacy	131	219	+67
Supplies	94	112	+19
Diagnostics and Treatment	1014 (51.7)*	1661 (83.2)	+64
Hematology	49 (5.9)	65 (7.0)	+33
Microbiology	41 (3.6)	92 (8.2)	+124
General chemistry	277 (17.7)	406 (28.6)	+47
Cardiovascular	57 (3.1)	141 (4.1)	+147
Radiology	174 (5.1)	162 (7.5)	−7
Respiratory therapy and cardio-perfusion	239 (9.3)	621 (19.6)	+160
Miscellaneous	177 (7.0)	174 (8.2)	−2

TABLE 3. Distribution of Percentage of Tests Ordered by Orderer in Teaching and Private Units

Service Unit	Physician Group	Number of Patients	Per Cent of Tests Ordered by: Total	Attendee	Resident	Other
All patients	All	105	100	48	40	12
Teaching	All	40	100	23	69	9
	1	21	100	26	69	9
	2	19	100	18	71	12
Private	All	65	100	73	12	15
	1	20	100	75	13	12
	3	45	100	72	12	16

distribution of tests by order for patients in each unit. As expected, in the teaching unit residents order 69 per cent of all tests and attending physicians order 23 per cent. This contrasts with the private unit, where residents order only 12 per cent of all tests, while attending physicians order 73 per cent. The residents have supervised primary responsibility for the teaching unit and are allowed to write orders with later consultation on teaching rounds. On the private unit, they can write orders only with the attending physician's direct prior approval. In Table 3, group 1 physicians are those who have patients in both teaching and private units (same as in Table 2). Yet the data imply that although they are the attending physicians in both units, the physicians are much more involved in the care of their patients on the private unit as compared with the teaching unit.

To examine the pattern of involvement on a day-to-day basis, Figure 1 illustrates the proportional ordering pattern of diagnostic tests for the first nine days of hospital stay among teaching and private unit patients. The data clearly indicate that attending physicians' involvement is virtually complete among their own private unit patients, as is expected. Among teaching unit patients, attending physicians are involved much less frequently, as demonstrated by the proportion of diagnostic tests ordered, on no day being responsible for a majority of care as measured by the proportion of tests ordered. The residents are the

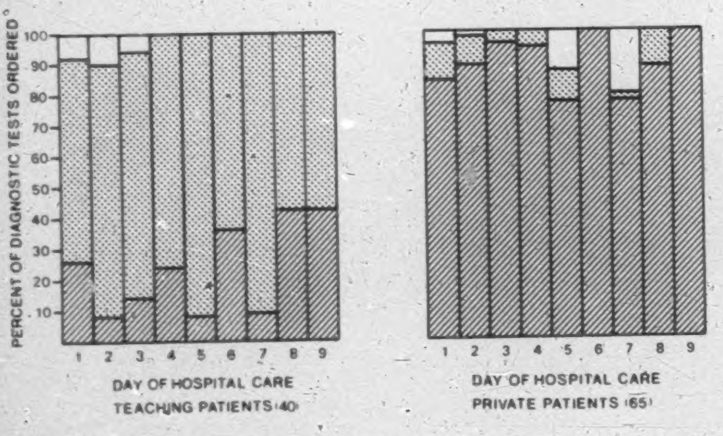


FIG. 1. Per cent distribution of diagnostic tests ordered by attending physician and resident; and standing orders for teaching and private service patients. (Striped area, attending physician; stippled area, resident; blank area, standing.)

primary communicators to patients and their families on the teaching service, whereas private attending physicians do most of the talking to patients and their families on the private service.

Figure 2 presents a comparison of ordering patterns for attending physicians (group 1 physicians from Table 3) who had patients in teaching as well as in private units. Here, also the same pattern persists, and the relationship of the attending physician in the private unit as much stronger. For private unit patients, attending physicians initiate virtually all orders, and residents play little role. Attending physicians in the teaching unit, however, are on no hospital day more involved with their patients than are the residents.

Discussion

The results of the present study raise several issues concerning reimbursement policy for graduate medical education: 1) reimbursement policy by third parties for resident salaries and for the portion of teaching physicians' time devoted to graduate medical education; 2) the level of involvement for physicians in dual roles when they are attending physicians for private patients, and simultaneously engaged in graduate medical education and supervising the patient care delivered by residents; and 3) the amount of extra utilization of hospital resources in teaching units.

Under the present method of reimbursement by third parties, the resident salaries and fringe benefits, and the part of teaching physicians' salaries for supervising graduate medical education are permitted to be included in the general operating expenses of a teaching hospital. The arrangement leads to a somewhat uniform increase in charges for each unit of service produced for patient care and subsequently billed to third parties. Thus, the charge for each unit of service includes a portion that relates directly to the teaching program. Under this arrangement every patient is paying for the cost of medical education, at a uniform rate. There is a charge differential between a hospital that is involved with a graduate medical education program and another that is not.

In a teaching unit, the care is primarily delivered by the resident under the supervision of a teaching physician. It is also apparent from two earlier studies[1,2] that the utilization of hospital services is greater in teaching units. In the present study, in spite of a control on primary diagnosis and on supervisory physicians, the utilization is 64 per cent higher in teaching units than in private units.

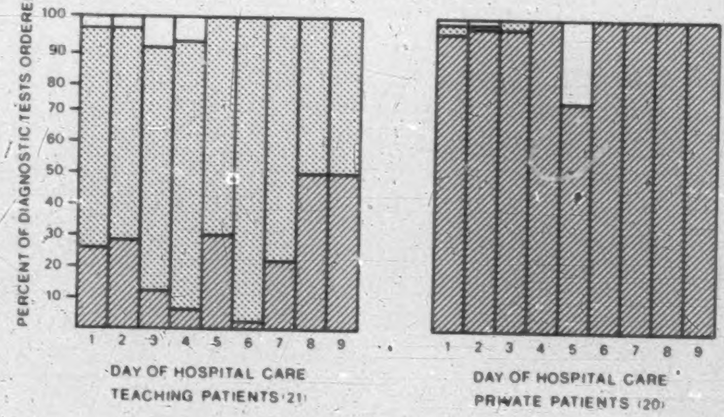


FIG. 2. Per cent distribution of diagnostic tests ordered by attending physician and resident, and standing orders for teaching and private service patients with same group of attending physicians. (Striped area, attending physician; stippled area, resident; blank area, standing.)

In two earlier studies[6,7] on the purpose of the tests ordered by residents,[6,7] it was found that approximately 10 per cent of the tests were ordered for defensive and educational purposes. Of course, residents are likely to order extra tests because they are learning medicine, and, as their experience increases, they are likely to be more careful in ordering.

The increase in number of diagnostic tests brings with it another increase in interpretation fees for excessive testing. For example, in the private unit, an average of 2.5 EKGs per patient were ordered, as compared with 3.6 in a teaching unit. Thus, the interpretation fee for 1.1 EKGs will be extra in a teaching unit.

The progressive increases in the cost of patient care for the three types of patients are shown in Figure 3. The figure illustrates the number of items that should be considered when estimating the total cost of graduate medical education and that were not meant to be quantified in the study. The first bar in Figure 3 illustrates that, in a nonteaching hospital, there are generally three components of patient charges for inpatient care: 1) hospital charges; 2) interpretation fees of support physicians, such as pathologists, radiologists and other consultants; and 3) fees of attending physicians. The second bar represents the private unit of a teaching hospital; there are added charges due to the existence of teaching programs. These charges include only the salaries and fringe benefits of the residents, and the portion of salary and fringe benefits of the teaching physicians. In the last bar, which represents the teaching unit of the teaching hospital, there are two additional items that increase the charges as compared with those of a private unit. The items are overutilization of diagnostic tests and procedures, and additional reimbursement for services of support physician services, such as pathology, radiology and cardiology, as a consequence of overutilization.

The study uncovers two key dilemmas in medical education. The first dilemma is documented by the excessive utilization and the teaching physician's salary added to the charges in the teaching unit. The second derives from the basic process by which resident education in hospitals is financed.

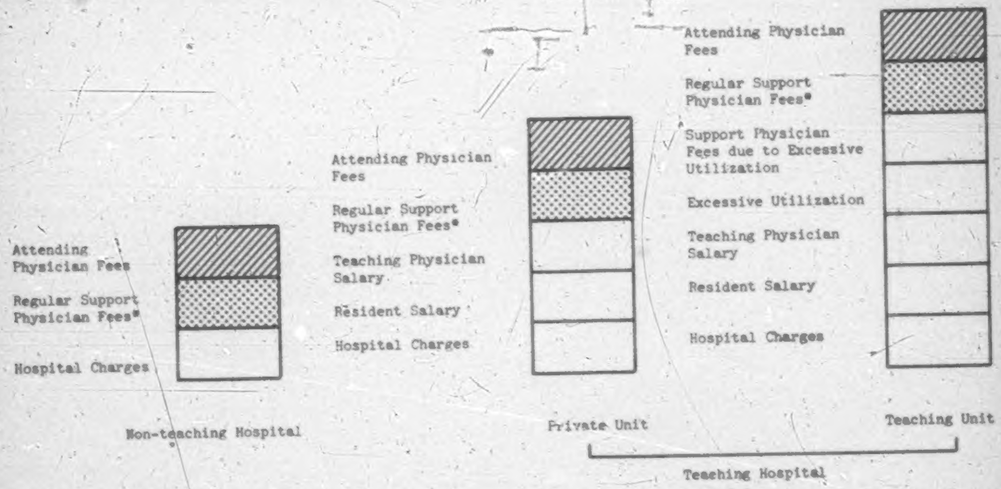


FIG. 3. Progress increase in charges for patients with congestive heart disease as a consequence of graduate medical education.

* Interpretation fees of pathologists, radiologists, cardiologists, etc.

Two methods of education are used in current residency training programs: one method utilizes learning by doing under guidance and supervision, whereas the other represents learning from a role model. A difference is seen in the levels of control exercised by teaching physicians over patient care on the private unit compared with the teaching unit. When teaching physicians exercise close control of patient care, the cost of the health care is clearly lower. Two levels of health care cost, if not of health care quality, exist in many teaching hospitals, and the difference raises questions of ethics and public policy. For example, in 1976 there were 15,381 full-time physicians in hospitals belonging to the Council of Teaching Hospitals[8] and 31,342 in other associated hospitals out of a total of 343,876 professionally active physicians. As the present analysis indicates, the level of involvement in patient care is different, but the level of reimbursement is the same. The policy is based on the premise that teaching physicians should be reimbursed equally if "personal and identifiable"[9] services were provided. However, the present data indicate the contrary. Therefore, the reimbursement policy, as presently practiced, ignores the basic difference in the level of involvement. Obviously, if the reimbursement is appropriate for one group, then it is inappropriate for the other.

Since the cost of teaching physicians' salaries is built into hospital overhead, and subsequently into daily bed rates, the problem of teaching physicians' salaries added to their fees as attending physicians bears reevaluation. There have been no studies evaluating the loss in fees for teaching physicians due to their teaching time. Teaching physicians may actually gain more in that they have "free help" to assist them in managing patients. Private physicians are reimbursed at the same rate but must manage patients with more of their personal time.

Three alternatives stand out as solutions to the problems posed. The first reflects a maintenance of the status quo, with a recognition that some costs of medical education are logical components of health care in general and should be borne by patients and their insurance companies. The slight percentage increase cost is small compared with that of other hospital charges, and the residency program allows attending physicians to be more efficient in other aspects of their practice.

The second solution represents a recognition of the need for closer supervision to prevent excessive utilization on the teaching unit and to bring health care costs to the level of the private unit. The amount of teaching physicians' salaries would depend on the recognition of the importance of teaching time and the need to attract teachers to the teaching hospital, as well as on the increased quality of care resulting from a teaching environment.

The third solution recognizes each individual component of health care costs and supports only those essential to patient care and physicians' services to be covered by insurance or private pay. Teaching functions would be supported by community, governmental or university educational funds, and would reflect the public recognition of the need for medical education and a willingness to support it. Whereas this solution would make health care and medical education more sensitive to public need, it would also make health care and medical education sensitive to public whim and political pressure.

In many teaching hospitals, attending physicians do not receive salaries from the hospital. A dilemma arises when full-time, hospital-based physicians on salary have fees collected for them by the hospital. No dilemma exists if their salaries are totally derived by fees collected and are not reflected in hospital charges. However, there is often an incentive for physicians to obtain bonuses beyond their salaries by increasing their patient contacts and, therefore, their patient billing. This can easily be done by increasing their patient contacts through the use of residents and then billing for the resident contacts.

The study implies that as more primary responsibility is given to residents on a teaching unit, greater costs result by virtue of the residents' inexperience and uncertainty. It can be surmised that close supervision and experience improve physicians' efficiency and decrease excessive utilization. Increased supervisory responsibility of teaching physicians could produce the close supervision and exposure to experience, but requires more of the teaching physicians' time and, therefore, more dependance on salary from the teaching hospital. Whether the increased efficiency and cost savings produced by having more full-time teaching faculty in hospitals balance the cost of salaries needs to be studied.

The ultimate solution, it is hoped, depends on decisions made by informed interaction between teaching hospitals, organized medicine, medical schools, and consumers and their representatives. The reevaluation of traditional approaches, advantages and disadvantages of different systems, and the basic goals of residency programs need to be considered before drastic changes are made. Some basic research and comparative studies must be carried out to provide the information needed for informed decisions.

Finally, from the standpoint of hospital income, the factors described in the study compound and increase one another, as long as the system remains openended. As limits are placed on hospital rate increases, more pressure is placed on the hospital to eliminate teaching physicians' salaries and to encourage fee collection by current separate mechanisms. Indeed, a clear separation of physicians' fees and the hospital bill would allow for more clear delineation of cost factors. Moreover, if residents assist physicians in their work, then their salaries should be derived as a component of the physicians' fees, and not the overall hospital charges. Medical care and liability consequences would be clearly the ultimate responsibility of attending physicians. It is hoped that attending physicians would have incentive for greater supervision and improved role modeling. The possibility that attending physicians would become dictatorial would be balanced by the competition among hospitals for a limited number of residents and the need to provide a good teaching environment to attract residents.

References

1. Schroeder SA, O'Leary DS. Difference in laboratory use and length of stay between university and community hospitals. J Med Educ 1977;52:418.
2. Martz EW, Ptakowski R. Educational costs to hospitalized patients. J Med Educ 1978;53:383.
3. Department of Health, Education and Welfare. Interim Report of the Graduate Medical Education National Advisory Committee to the Secretary. Washington, D.C.: Department of Health, Education and Welfare, April 1979.
4. Medicare–Medicaid Reimbursement Policies. Institute of Medicine of the National Academy of Sciences, March 1979.
5. New York Heart Association Criteria Committee. Nomenclature and Criteria for the Diagnosis of Diseases of the Heart and Great Vessels. New York: Little, Brown, 1973.
6. Garg ML, Gliebe WA, Elkhatib M. The extent of defensive medicine: some empirical evidence. J Legal Med 1978;6(2):25.
7. Garg ML, Gliebe WA, Elkhatib M. Medical practice for educational purposes: the internal medicine resident in a hospital 1978;52(4):97.
8. COTH Director. Educational Programs and Services. Council of Teaching Hospitals, Association of American Medical Colleges, 1978.
9. Hadley J. Medical Education Financing: Issues and Options. Department of Health, Education and Welfare. [Publication No. (PHS) Washington, D.C.: 79-3261.]

MEDICAL CARE
July 1982, Vol. XX, No. 7

Effectiveness of an Ambulatory Care Telephone Service in Reducing Drop-In Visits and Improving Satisfaction With Care

CLARENCE F. STIREWALT,* MARGARET W. LINN, PH.D.,† GUSTAVO GODOY,‡
FELIPE KNOPKA, M.D.,§, AND BERNARD S. LINN, M.D.‖

The purpose of this research was to determine whether a specialized telephone service could cut down on unscheduled visits to ambulatory care and improve satisfaction with care. Patients who were to receive care (N = 561) were assigned randomly to a specialized telephone service or a control group. They were measured with regard to satisfaction with care before being assigned and 6 months later. The number of scheduled and unscheduled visits also were monitored for a 1-year follow-up period. The telephone service, which provided a triage system for referrals as well as a source of central contact for scheduling or for complaints, helped to improve satisfaction and reduce the number of unscheduled visits within 6 months. Such a service also may have application in other problem areas in the delivery of quality ambulatory care.

TWO EXTREMES of patient behavior have created problems in the delivery of ambulatory care. At one extreme, there are the "no-show" patients who leave holes in appointment scheduling so that the most cost-effective use of the care providers' time is diminished. At the other extreme, there are the "drop-in" patients who disrupt scheduling in the opposite manner, overloading the system so that the quality of services provided may be jeopardized. Unfortunately, these two types of patients seldom balance each other out, so that "drop-ins" appear conveniently when "no-shows" occur. Thus, the problem is compounded.

Over a number of years, in an extremely busy outpatient department, the number of patients who came without appointments had escalated to such an extent that a full-time physician had to be assigned each day just to manage these patients. A preliminary study comparing the drop-in population with clinic users who usually sought care by appointment indicated less satisfaction with care among drop-in patients. Furthermore, the drop-in patients tended

---

From the Veterans Administration Medical Center and the University of Miami School of Medicine, Miami, Florida.

* Chief, Medical Administration Service, VA Medical Center.

† Director, Social Science Research, VA Medical Center, and Professor of Psychiatry, University of Miami School of Medicine.

‡ Administrative Officer, Research Service, VA Medical Center.

§ Associate Chief of Staff for Ambulatory Care, VA Medical Center, and Assistant Professor of Medicine, University of Miami School of Medicine.

‖ Associate Chief of Staff for Education, VA Medical Center, and Professor of Surgery, University of Miami School of Medicine.

Supported by VA Health Services Grant 310-2.

Address for reprints: Margaret W. Linn, Ph.D. (151), VA Medical Center, 1201 N.W. 16th Street, Miami, FL 33125.

0025-7090/82/0700/0739/$01.00 © J. B. Lippincott Co.

to have a history of high clinic utilization with repeated drop-in behavior patterns.

This paper describes a prospective randomized trial of a specialized telephone system designed to cut down on drop-in visits and improve patient satisfaction with care.

Several studies have determined that symptoms and ability to pay are the best predictors of the use of health services. Anderson[1] found these two factors more important than cultural and social factors. Bice and White[2] showed that perceived symptoms were the primary variable explaining utilization, not only in this country but in several other countries as well. Avnet[3] found no difference in utilization due to income when there was no financial barrier, such as when care was provided through comprehensive insurance programs. Greenlick[4] made the same observations in a study of regular subscribers to the Kaiser Foundation Plan. Likewise, Bice et al.,[5] Mechanic,[6] and Montiero[7] concluded that sociocultural differences in utilization were diminished greatly when capacity to pay for services was equalized.

The present study concerns ambulatory care provided under the auspices of the Veterans Administration, where the ability to pay was not seen as an important variable in explaining utilization, either by drop-in or scheduled appointments. The person's perception of health, however, was seen as a major factor in determining utilization. In fact, the rationale for selecting a specialized telephone system for testing was based on the assumption that it was just this perception by the patient of the need to seek immediate medical care that resulted in an inability to wait for a scheduled appointment.

Weinerman[8] believed that patient satisfaction with care influenced the degree to which medical care was sought and the extent to which advice was accepted and heeded. Stoeckle et al.[9] suggested using satisfaction data to plan and organize facilities, and to set priorities for medical education and public health programs. Several investigators[10–12] have concluded that patient satisfaction helps in understanding why people do or do not seek care. Patient acceptance often looms as a potential barrier to implementating new modes of delivering services. Donabedian[13] suggested that patient satisfaction, along with health status, was the ultimate outcome in evaluating quality of medical care. Again, the idea of a telephone service which would provide immediate access to a health system for patients who seemed to need and seek fast access to care, appeared to have the potential to influence the degree of satisfaction, which might in turn lead to modified behavior in seeking appointments for care.

The telephone's role and potential in health care has not been explored very thoroughly. In 1971, 13 per cent of patient contacts with physicians were provided by telephone.[14] Delegation of telephone responsibility to nonphysicians has been acceptable to patients.[15] Brown[16] and Ott[17] found deficiencies in history taking and accuracy of advice among house officers and health associates who were not specially trained in telephone care. However, Strain[18] found few errors and a high level of parental acceptance with a nurse who trained for this role. In special programs for chronic illness and suicide prevention, nonphysicians have been trained as primary care providers.[19–21] The telephone has been used in poison control centers as an aid in diagnosis and in offering medical consultation to rural areas.[22] Telephone service to obtain results from throat cultures and triage by telephone in emergency room care also have been described.[23,24]

The purpose of this study was to evaluate the effectiveness of a specialized telephone service in ambulatory care using a prospective controlled study for patients who dropped in for care without an appointment in regard to improving patient satisfaction with care and in reducing the number of drop-in visits.

Method

The Setting

The Miami Veterans Administration Medical Center has one of the largest outpatient loads among VA Medical Centers. Over 200,000 outpatient visits are scheduled to its 38 clinics per year. Computerized scheduling and preordering of medical records had improved the efficiency of services for patients with scheduled appointments. Patients who dropped in for their care, however, disrupted the system. Furthermore, patients had to wait while their records were being located. They also lost continuity of care through being seen by a different physician on each visit. Each day, one physician was assigned to see drop-in patients only. It was known from the pilot study that these patients often wanted answers to questions or help with problems that did not require the services of a physician. Pharmacy services, social work, dietetics, nursing, medical administration, and other services frequently were a more appropriate source of contact than the physician. Few problems actually appeared emergent enough to require immediate on-site medical attention. It seemed likely that with a telephone service patients might be influenced to schedule an appointment if a clinic visit were needed instead of just dropping in.

The Telephone Service

A clerk trained thoroughly in clinic scheduling and who was familiar with services in the hospital and outpatient department was selected to provide the telephone contacts. Protocols for telephone triage and an extensive referral network were developed for handling patients' problems. First, the clerk determined why the patient was calling. If, for example, the patient presented family problems, a referral was made to social work service. If the patient was out of medication and was eligible for a refill, the pharmacy was contacted. The same triage was followed in regard to general medical, nursing, prosthetic or medical record services. Questions dealing with eligibility, scheduling of appointments, complaints or simply the need to talk with someone were handled directly by the telephone clerk. One of the functions of the clerk was to explain services available on a scheduled basis and to offer one source where patients could call with their questions or needs. When referral was made to other services, a form was sent to that service to follow-up on their action so that the final disposition was known. If a patient was assigned to the experimental group, and therefore eligible to use the telephone service, he was introduced to the telephone operator and given a card that included the patient's name, unique number and the telephone service number. Calls were transcribed and coded for length, caller, reason for call, referral and outcome. The telephone service operated each day during clinic hours from 8:00 to 4:30.

The Research

A multidisciplinary team of investigators planned the study and met monthly to monitor its progress. Data on characteristics of patients entered into the study and types of referrals were reviewed. Randomly selected tapes of calls also were reviewed. Intake and follow-up statistics with reasons for patient loss were monitored as well. Statistical analyses were not done until the end of the study so that bias would not be entered from knowing results of unblinded comparisons.

A research assistant randomly selected patients from among those who came in for their care without appointments over a 2-year period of intake (1979-80). During these 2 years, approximately 7500 drop-in visits occurred among 3160 different patients. The method used for selection into the study was to screen drop-in patients during 2 hours of each day. The screening hours were rotated so that different hours were covered each day of a week, with the next week rotating to different hours for the days, so that within a 4-week cycle, all hours of each day had been covered. Male patients who met the criterion of living within the area covered by the clinic were eligible for inclusion in the study if they agreed to sign informed consent. Eight per cent of those selected did not agree to participate. Patients who signed informed consent were interviewed concerning demographic characteristics such as age, education, race, social class (measured with the Hollingshead two-factor index), household composition, marital status, presenting symptoms, pattern of use of medical services, distance from the clinic and self-perception of health (measured on a 5-point scale).

Patients then completed two satisfaction-with-care scales. A 45-item scale[25] that measures three areas of satisfaction was used. These areas are professional competence of physicians, personal qualities of physicians and cost/convenience of care. Statements such as "You may have to wait a little, but you can always get a doctor," were rated on 5-point strongly-agree to strongly-disagree scales. In addition, ten items that measure a more general satisfaction with care were used. The latter scale represents the super-factors of the VA Satisfaction with Care Scale, which is scored on 4-point scales.

After baseline testing, patients assigned randomly to an experimental group (i.e., eligible to use the telephone) or a control group (i.e., handled in the usual way) by their opening a sealed envelope which contained the group assignment (made with a table of random numbers). The experimental patients were given the telephone card and encouraged to call, rather than drop in, for care. Both experimental and control patients were advised that they would be contacted again in 6 months. Data were collected from their medical records to describe their diagnoses, number of drop-in visits, scheduled visits, hospitalizations during the prior 6 months and number of medications.

Six months after baseline testing, groups were retested on the same two satisfaction-with-care scales and the same information collected from their medical records describing use of scheduled and unscheduled visits and hospitalizations. An additional record follow-up was done at 12 months. Thus, the study was completed in July 1981. Data also were collected on perception of health, employment hours and overall evaluation of care. Patients in the experimental group continued to use the telephone service for as long as the study was in progress, and scheduled and unscheduled visits were recorded at 6 and 12 months for experimental and control patients.

Data Analysis

The primary outcomes of the study were satisfaction with care and number of drop-in visits. The hypothesis was that the telephone service would be effective in increasing satisfaction with care and reducing the number of drop-in visits. Data were analyzed by comparison of the outcomes of the experimental and control groups. For satisfaction-with-care ratings, the baseline scores were held constant comparing satisfaction at 6 months in a multivariate analysis of covariance as recommended by Bancroft[26] for pre–post randomized designs. Number of drop-in visits over the follow-up of 6 and 12 months was analyzed by standard analysis-of-variance techniques.

Results

The Sample

A total of 561 patients were studied, with 279 randomly assigned to the experimental group and 282 to the control group. A total of 82 per cent were successfully followed for their reevaluations concerning satisfaction with care, and all but 9 per cent were followed for determination of scheduled and unscheduled visits. The reasons for loss of data were death, moving out of the area and inability to locate patients. Comparison of baseline characteristics between those lost to follow-up and those on whom all data were complete showed no significant differences. Thus, the sample seemed to be representative of all who entered the study.

Table 1 shows the background characteristics of experimental and control patients. There were no significant differences on any of the variables describing personal characteristics. Patients averaged about 54 years of age; they had high school educations; they were lower middle class. Most were married, were residents of the area and lived about 22 miles from the clinic. They assessed their health as fair (3.2 average on a 5-point scale).

Furthermore, the experimental and control groups did not differ significantly on characteristics that described their use of services over the 6 months before they entered the study. Table 2 shows that each group averaged 2.92 and 2.46 drop-in visits. Slightly over half of their total visits were drop-in visits. Groups averaged about four diagnoses, the major ones being hypertension, diabetes, arteriosclerotic heart disease, emphysema, peptic ulcer, arthritis and skin disorders. Each group received about three medications. The control group had slightly more (although not statistically significant) hospitalizations than the experimental group.

Use of the Telephone Service

An average of 3.3 calls per year per patient in the experimental group was recorded, 86 per cent of the group using the telephone service at some time. Calls were received throughout the day on a rather even basis, with no particular hour showing higher use than any other. The duration of each call averaged 7.2 minutes (standard deviation of 7.6). Almost half of the calls (49 per cent) were handled directly by the telephone operator, the remainder (51 per cent) being referred to other sources. The most frequent referrals were for questions about the person's medical condition, which usually were referred to a nurse or physician (33 per cent). Some patients asked for appointments or when there seemed to be a need, appointments were

TABLE 1. Comparison of Backgrounds Between Experimental and Control Patients

Variables	Groups: Experimental	Groups: Control	F-Ratio
Age	54.05	53.13	0.47
Education (years)	12.23	12.08	0.23
Race (white)*	0.84	0.83	0.02
Social class	3.75	3.84	1.18
Married*	0.85	0.86	0.01
Total in home	1.36	1.39	0.07
Miles from clinic	21.43	22.20	0.17
Florida resident*	0.97	0.97	0.11
Self-assessed health	3.26	3.24	0.12

* Items can be read as per cent of group by disregarding the decimal point. Social class was measured 1–5 with 5 being the lowest social class. Self-assessed health was rated 1–5 with 5 being the poorest.

TABLE 2. Comparison of Experimental and Control Patients on Health Variables at Baseline

Variables	Groups		F-Ratio
	Experimental	Control	
Average number over past 6 months			
Drop-in visits	2.92	2.46	1.06
Scheduled visits	1.79	2.18	1.33
Hospitalizations	0.17	0.25	2.02
Current number			
Medications	3.26	3.09	0.78
Diagnoses	4.67	4.35	1.28

scheduled (26 per cent of the calls). About 12 per cent of the calls were referred to the pharmacy, and less than 10 per cent for each of the following: general information, cancelling appointments, complaints, questions about eligibility for care, admission to the hospital, obtaining medical records, travel to the clinic, prosthetics, transferring records and help with completion of medical forms.

Satisfaction with Care

Table 3 shows the adjusted and unadjusted mean scores at 6 months on the two satisfaction-with-care scales. There were no differences between experimental and control groups at baseline testing. At 6 months, with prescores covaried, experimental patients had more favorable scores on all three factors of the 45-item scale, with professional competency, personal qualities, and cost/convenience of care differing between the two groups at the $p < 0.05$ level for each factor, the overall multivariate level being statistically significant ($p < 0.05$) for the three factors together.

On the 10 items that represented the super-factors of the VA Satisfaction with

TABLE 3. Comparison of Experimental and Control Patients on Satisfaction with Care

Variables	Unadjusted Postscores		Adjusted Postscores		F-Ratios	
	Experimental	Control	Experimental	Control	Unadjusted	Adjusted
Three-factor scores for satisfaction with:						
Professional competence	36.03	36.99	36.02	7.13	1.63	3.26*
Personal qualities	35.15	36.71	35.15	36.56	2.04	2.89*
Cost/convenience of care	42.85	43.74	41.87	43.99	2.99*	3.78*
Multivariate F (3 scores)						1.45*
Ten-item scale for satisfaction with:						
Overall care	3.29	3.07	3.32	2.90	4.10*	4.99*
Promptness of care	2.92	2.74	3.33	2.51	3.70*	3.98*
Consideration	3.53	3.31	3.64	3.25	3.00*	3.40*
Multivariate F (10 items)						1.79*

* $p < 0.05$.
Higher scores on three-factor scores are less favorable responses.
Higher scores on the 10 items are more favorable responses.

Care Scale, three were significantly different at 6 months at the 0.05 level. Again, the multivariate level for all 10 items was significant in discriminating the two groups at the 0.05 level. The significant item discriminators, in favor of the experimental group, were overall satisfaction with care, promptness and consideration. Although the remaining items did not reach significant levels for differences between the groups, they were all in the direction of the experimental group being more satisfied with care at 6 months.

Scheduled and Unscheduled Visits

Table 4 provides data on the number of scheduled and unscheduled visits made during the follow-up of 6 and 12 months. At 6 months, the number of drop-in visits was significantly lower in the experimental group (1.61 average) as compared with the control group average of 2.11. This difference was at a 5 per cent level of significance. The number of scheduled visits did not differ significantly between experimental and control patients. However, at 12 months, the number of drop-in visits did not differ between the two groups, nor did the number of scheduled visits. The total number of visits averaged about 3.8 for both groups. However, the number of hospitalizations was significantly higher in the control group ($p < 0.05$) than in the experimental group and could have accounted for their being in the community less time during follow-up and therefore not as likely to drop in as much. Diagnoses and medications were not different between groups throughout the study.

Discussion

The hypothesis that the telephone service would improve satisfaction with care was supported. After 6 months exposure to a specialized telephone service, experimental patients had significantly more positive attitudes as measured by the three factors on the Satisfaction with Physicians and Primary Care Scale as well as on the more general 10-item Satisfaction with Care Scale. More satisfied ratings on the cost/convenience of care are easily under-

TABLE 4. Comparison of Experimental and Control Patients at Six and 12 Months Follow-up on Health Variables

Variables	Groups		F-Ratio
	Experimental	Control	
Six Months Follow-up			
Average number over past 6 months			
Drop-in visits	1.61	2.11	3.63*
Scheduled visits	2.64	2.54	1.26
Hospitalizations	0.19	0.33	2.60
Medications	3.02	3.11	0.03
Diagnoses	4.82	4.23	1.54
12 Months Follow-up			
Average number over past 6 months			
Drop-in visits	1.65	1.35	1.58
Scheduled visits	2.18	2.24	0.66
Hospitalizations	0.16	0.41	3.12*
Medications	3.12	3.26	0.14
Diagnoses	4.38	4.46	1.03

* $p < 0.05$.

stood, because the telephone service was designed to make care more convenient and access more easily obtained. Increased satisfaction in regard to the professional competence and personal qualities of the physician is less easily understood, since physicians did not operate the telephone service, nor was any intervention in the doctor–patient relationship attempted. One answer could be that the scale, although purported to measure three discrete dimensions, measured a more generalized satisfaction with care, in reality. This could be supported by the moderately high intercorrelations between the three factors, which ranged from $r = 0.46$ to $r = 0.55$.

On the other hand, more drop-in visits were correlated with more negative attitudes on all three factors at about the same moderate levels of correlation. Therefore, when the experimental group had fewer drop-in visits as a result of the telephone service, then satisfaction could be expected to improve in all of these areas. The reason that more negative attitudes toward professional competence and personal qualities of physicians were associated with drop-in visits may be from lack of continuity of care when patients come without appointments. It may be that physicians assigned to cover drop-in visits were less satisfied when they did not see their own patients and this dissatisfaction may have been transmitted to the patients, thus influencing the patient's attitude. For whatever reason, all three factors scores concerning satisfaction with care were improved significantly in experimental patients, compared with controls. Improvement is further confirmed by the use of another scale, where the overall multivariate difference favored experimental patients more than controls. Items that contributed most to discrimination between the groups (overall care, promptness and consideration) seemed highly relevant to the services provided by specialized telephone service.

The hypothesis that a telephone service could reduce the number of drop-in visits was proven after 6 months but not at 12 months. Drop-in visits went from a baseline rate of 2.9 in the experimental group to 1.6 at 6 months. The drop-in rate stayed about the same in control patients during the same time (2.4 to 2.1). Number of scheduled visits did not differ significantly between the experimental and control patients at baseline or 6 months; however, the trend was toward an increase in the number of scheduled visits for experimental patients at 6 months. This would seem to be a logical finding since one of the functions of the telephone service was to schedule visits for patients when it was determined that a visit was necessary. Thus, the total number of visits was similar for both groups with the difference being in the way they utilized services, by appointment or not. One could argue, that utilization of care was not affected by the telephone service, even though drop-in visits were reduced at 6 months. One answer is that quality of care, as measured by patient satisfaction, was improved when drop-in visits were reduced. Certainly, it is a more efficient delivery of health services when appointments are scheduled in advance. It also may be less costly. Cost will be addressed in another article.

Patients were followed in regard to their utilization patterns through a 6-month period of time. By 12 months, drop-in visits were not significantly different between experimental and control subjects. The number of drop-in visits remained about 1.6 for experimental patients and had dropped to 1.4 in controls. What appears to have occurred is that both the rate of drop-in visits and hospitalization had stabilized in experimental patients over time. What happened with control subjects is that hospitalization rates increased over time, even though control subjects did not differ significantly from experimental patients in diagnoses (kind or number) and number of medications. If diagnoses and medications can be used as indicators of severity of condition, then control subjects who were similar to experimentals were using the hospital more by 12 months. If they used the hospital more, then they had less time in the community to use out-patient services. More importantly, it has been suggested that if outpatient visits were used more effectively, then rates of hospitalization might decrease. Unfortunately, this seldom has been found in studies.[27] Perhaps patients who utilize scheduled appointsment with the same physician have less of a need for hospitalization. Furthermore, patients who are more satisfied with their care may not need additional care through hospital admission as frequently.

Overall, a specialized telephone service for ambulatory care patients was effective in improving satisfaction and reducing drop-in visits. It also may be effective over a longer follow-up in reducing hospitalization as well. Patient response to the telephone service was favorable, with about three fourths of the patients who were eligible calling an average of three times. From another standpoint, even though the telephone was effective in reducing drop-in visits, it did not eradicate the drop-in problem completely. Even after 12 months, patients in the experimental group still averaged slightly over one drop-in visit during a 6-month period of time. Some of these drop-in visits came from the small proportion of the sample who did not use the telephone at any time during the study. The fact that a large proportion of the calls could be handled by nonphysicians suggests that a telephone may be an effective and efficient means of providing service, so that physician time can be used more effectively. Problems encountered in scheduling outpatient visits are not unique to our center; therefore, a telephone service may have wide application not only for drop-in patients but for other problem areas in delivering high quality ambulatory care.

References

1. Andersen R. A behavioral model of families' use of health services. Chicago: University of Chicago, Center for Health Administration Studies, Research Series 25, 1968.

2. Bice TW, White KL. Factors related to the use of health services: an international comparative study. Med Care 1969;7:124.

3. Avnet HD. Physician service patterns and illness rates. New York: Group Health Insurance, Inc, 1967.

4. Greenlick MR, Freeborn DK, Colombo TJ, Prussin JA, Saward EW. Comparing the use of medical care services by a medically indigent and a general membership population in comprehensive prepaid group practice program. Med Care 1972;10:187.

5. Bice TW, Eichhorn R, Fox PD. Socioeconomic status and use of physician services: a reconsideration. Med Care 1972;10:261.

6. Mechanic D. Illness and cure. In: Dosa J, Antonavsky A, Zola IK, eds. Poverty and health. Cambridge: Harvard University Press, 1969:191.

7. Montiero L. Expense is no object . . . : income and physician visits reconsidered. J Health Soc Behav 1973;14:99.

8. Weinerman ER. Patients' perceptions of group medical care. Am J Public Health 1964;54:880.

9. Stoeclke JI, Zola IK, Davidson GE. On going to see the doctor: the contribution of the patient to the decision to seek medical aid. J Chronic Dis 1963;16:975.

10. Coe RM, Wessen AF. Social-psychological factors influencing the use of community health resources. Am J Public Health 1965;55:1024.

11. McKinlay JB. Some approaches and problems in the study of the use of services: an overview. J Health Soc Behav 1972;13:115.

12. Lewis CE. Acceptance of physicians' assistants. Hospitals 1971;45:62.

13. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966;44:166.

14. U.S. Department of Health, Education, and Welfare, Volume of Physician visits, 1966, 1967, 1969 and 1971, Public Health Service Publication No. 49, 75 Series 10, Number 49. Washington: U.S. Government Printing Office.

15. Patterson P, Bergman A, Wedgewood R. Parent reaction to the concept of pediatric assistants. Pediatrics 1969;44:69.

16. Brown SB, Eberle BJ. Use of telephone by pediatric house staff—a technique for pediatric care not taught. J Pediatr 1974;84:117.

17. Ott JE, Bellaire J, Machota P, et al. Patient management by telephone by child health associates and pediatric house officers. J Med Educ 1974;49:596.

18. Strain J, Miller J. The preparation, utilization and evaluation of a registered nurse trained to give telephone advice in a private pediatric office. Pediatrics 1971;47:1051.

19. Wang MK, Kirkham FT. Use the telephone in a health maintenance service for the chronically ill—a preliminary report. J Chronic Dis 1971;24:489.

20. Williamson JW, Goldberg E, Packard M. Use of simulated patients in evaluating patient management skills of telephone counselors, a proposal. Crisis Intervention and Counseling by Telephone.

21. Littman RE, Farberow NL, Shneidman ES, et al. Suicide-prevention telephone service. JAMA 1965;192:107.

22. Mapes RW, Lewis EW, Covell DG, et al. Feasibility study of pediatric telephone consultation service. Pediatrics 1972;50:307.

23. Grebin B, Asnes R. Dial-a-culture—a new application of the telephone to increase compliance. Abstracts of the Ambulatory Pediatric Association 1975;159.

24. Lattimer A. Telephone triage ("Totline") by nurses in a pediatric acute care clinic. Abstracts of the Ambulatory Pediatric Association 1975;188.

25. Hulka BS, Zyzanski SJ, Cassel JC, Thompson SJ. Scale for the measurement of attitudes toward physicians and primary medical care. Med Care 1970;8:429.

26. Bancroft TA. Topics in intermediate statistical methods. Ames, Iowa: Iowa State University Press, 1968.

27. Chassin MR. The containment of hospital costs: a strategic assessment. Med Care 1978;16(Suppl).

MEDICAL CARE
July 1982, Vol. XX, No. 7

Book Reviews

Antitrust and the Health Care Provider, by Martin J. Thompson, Germantown, Maryland, Aspen Systems Corporation, 1979, 250 pp., $30.00 cloth.

APPLICATION OF ANTITRUST LAW to the activities of health care providers has sent shock waves throughout the industry. Professional associations have been particularly reluctant to accept the fact that they are subject to the strictures on anticompetitive practices laid down in the Sherman, Clayton and Federal Trade Commission Acts. Martin Thompson's book comes at a time when health care providers are concerned about the antitrust implications of their collective actions. Unfortunately, the book is written primarily for lawyers and may not be geared to the audience most in need of developing a better understanding of what antitrust enforcement means in this market.

The book is divided into nine chapters, each of which, according to the author, was written to stand on its own. With the exception of the introductory chapter—a compact review of the antitrust statutes, jurisdictional issues and general implications of the antitrust laws for the health care market—each chapter discusses a discrete set of activities. The chapters vary considerably in the depth and quality of the underlying research, although they are consistently weak on including references to economic and social science research findings. Some chapters are good summaries of the statutory and case law, but overall the book lacks analysis of the policy implications of current law for health care providers. Consequently, it reads more like a hornbook for the fledgling attorney than a general treatise.

Part of the problem with this book is that it deals with an area of the law that is developing rapidly. It was written almost three years ago, and many of the legal issues covered in the book have now been litigated and decided by the courts. In fairness to the author, the merits of the book should not be judged solely on how well it anticipated subsequent events but instead on its intrinsic value.

Certain chapters of this book are well worth reading. In particular, chapter 3 ("Health Planning and other Government Interventions"), chapter 5 ("The Role of Trade Associations") and chapter 7 ("Relations with Health Insurers") represent decent summaries of the legal implications of the antitrust laws in these areas of provider activity. In chapter 3, the author reviews First Amendment protection for legitimate efforts by health care providers to influence governmental policy decisions, and other immunities from antitrust suits conveyed by the implied repeal and state action doctrines. Since these immunities are important to the disposition of antitrust cases challenging certain activities, Thompson's discussion of them this early in his book was a wise choice.

Chapter 5 is perhaps the best 40 pages in the book. Thompson briefly runs through the role of professional associations in setting standards through accreditation and certification, organizing group boycotts of nonmedical providers, conducting disciplinary proceedings, conducting boycotts of insurers and HMOs, setting fees, organizing peer review activities and lobbying. The only jarring inconsistency in this chapter is the inclusion of cooperative buying and selling arrangements in the sweep of professional association conduct. These activities deserve far more consideration than the two-page discussion here; ideally, they belong in a separate chapter along with similar arrangements used by hospitals or other health care institutions.

Chapter 7 is a concise little chapter that deals with provider/insurer relations—a topic of considerable importance in this industry. Thompson begins with a detailed analysis of the McCarran–Ferguson Act (which exempts the "business of insurance" from antitrust scrutiny), and concludes with a review of insurance boycotts, coercion and intimidation. It is in this latter area that antitrust enforcement is likely to intensify in the future as insurers attempt to hold the line on costs with providers. In fact, the Federal Trade Commission and the Department of Justice have brought several cases against providers who allegedly organized such boycotts. But the author does not share his views on where the case law appears to be going; instead, he merely recites current law. The result is a wooden cross-sectional look at the status quo, with no consideration of what that look portends for the future.

The remaining chapters are pedestrian restatements of the antitrust laws. Chapter 4, "Mergers, Acquisitions, and Affiliations," reviews an area of the law that, until recently, has had little application to health care institutions. Only since the rapid growth of hospital and

/

C5170-1

Medical Care

August 1982
Vol. 20 No. 8
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

A Review of Methods for Ambulatory Medical Care Evaluations

An Examination of the Factors That Influence Patient Referral Decisions

Use of Illness-Related Ambulatory Physician Services in Finland

Visits to Physicians by Asian/Pacific Americans

Clinical Evaluation vs. Economic Evaluation: The Case of a New Drug

Drug Product Selection in Illinois: The First Year's Experience

Evaluation of a Sequence of Diagnostic Tests Using the Workup of Ureteral Stone as a Model

The Use of Ancillary Services: A Role for Utilization Review?

Podiatric Surgery: The Need for a Second Opinion

Utilization of an Automated Multiphasic Health Testing System for Performing Prehospitalization Examinations

PUBLISHED BY

MEDICAL CARE
August 1982, Vol. XX, No. 8

An Examination of the Factors That Influence Patient Referral Decisions

ROBERT L. LUDKE, PH.D.

The results of a study to identify the factors that two types of referring physicians, general/family practitioners and general surgeons, consider when deciding whether or not to refer the patient, and where to refer the patient are presented and discussed. The factors for each type of decision were identified initially through interviews with physicians in active practice in Wisconsin and rank ordered according to their relative importance in making the referral decisions. These factors were then used in two alternative normative decision-making models to predict physician referral behavior. Factors related to the kind and quality of medical care the patient would receive appear to be most important in the decisions, with a high level of agreement between and among general/family practitioners and general surgeons as to the relative importance of the factors. Although the normative decision-making models correctly predicted a relatively high percentage of the referral decisions, not all of the factors were found to be important in explaining the behavior of physicians, and were not always consistent with the factors that were reported to be important.

AT SOME POINT during the process of delivering medical care to a patient, a physician may consider referring that patient to another physician or medical center. The decision to refer a patient may occur at any time during the care process, such as during diagnosis, after diagnosis but before treatment, or during treatment. The reason for referral may vary. It may be for diagnosis, confirmation of diagnosis, opinion regarding treatment, or treatment, and may depend upon the scope of practice of the referring physician, the diagnostic and treatment equipment and procedures available to the physician, time constraints, and a number of other potential factors related to the physician, the patient, and the referral physicians or centers. In contrast to consultation, where responsibility for a patient remains with the initial physician but advice or special studies from other physicians are requested, referral represents that act involved in transferring some or all responsibility for the patient's care, either temporarily or permanently, from one physician or health agency to another physician or health agency.[1-3]

Studies conducted in the United States found that physicians refer approximately 1–6 per cent of their patient visits, on the average, with large variation in referral rates among individual physicians.[1-11] Several concerns have been raised regarding the referral process. Clute,[12] Miller,[13] and Peterson et al.[14] have questioned whether appropriate referrals are being made. Others have raised questions about the adequacy of communication between

---

Graduate Program in Hospital and Health Administration, College of Medicine, The University of Iowa.

Supported by Grant Number 1 R03 HS 02351-0 from the National Center for Health Services Research, HRA.

Address for reprints: Robert L. Ludke, Ph.D., Graduate Program in Hospital and Health Administration, S-502 Westlawn, The University of Iowa, Iowa City, IA 52242.

0025-7079/82/0800/0782/$01.55 © J. B. Lippincott Co.

referring and consulting physicians.[1,6,11,15–17] Williams et al.[18] found failures on the part of consulting physicians to return patients to referring physicians.

Although previous research has provided some insight into the rates of referral relative to physician and patient demographic characteristics and into some of the major problems that appear to exist in the referral process, it contributes little to an understanding of why and how the process operates as it does. The study conducted by Shortell[3] is an exception, however. His model of physician referral behavior was based on social exchange theory. The basic premise was that physicians, occupying different levels of status within the medical community, perceive different rewards and costs associated with referring patients. Shortell concluded,

> In summary, between seven and 49 per cent of the variance in the referral rates is explained by exchange theory and related variables . . . neither the status index nor individual components of the index figures significantly in explaining the results . . . other variables such as board certification, organization of practice, caseload severity, and office visits per week explain much of the variance not accounted for by the status variable. (p. 69)

A subsequent study by Shortell and Vahovich[8] was conducted to address the hypothesis that patient-related variables are the most reliable predictors of referral rates for client-dependent physicians, such as general practitioners, and physician-related variables are the most reliable predictors of referral rates for colleague-dependent specialists, such as surgeons. Although they found that physician-related variables were the most important predictors of referral rates for surgeons, they also found that to be the case for general practitioners.

Given the frequency of referrals, the implications of the referral decisions and the potential concerns regarding the referral process, the referral process warrants further study. Particularly, greater insight into the variables which physicians take under consideration in making the referral decisions is needed. Through a greater understanding of the referral process, steps can be taken to begin assessing whether or not optimal use is being made of referring physicians, the referral process, and referral centers in terms of manpower, resources, and quality and cost of patient care. Also, insights can be gained into how and what modifications should and could be made in order to strengthen the referral system.

This paper takes a step in that direction by presenting and discussing the results of a study to identify the factors that physicians consider when faced with the possibility of referring a patient. The study focused on the referral of breast cancer patients after the referring physician conducted the initial history and physical examinations of the patient and performed such preliminary diagnostic procedures as chest x-ray, liver function tests, blood count and/or mammogram. Given this point in time in the patient's care process, the physician must decide whether or not to refer the patient for further diagnostic workup and/or treatment and if the decision is to refer the patient, where to refer that patient.

Since the factors that are considered may differ for each of these decisions, the study assessed the factors important for each decision. In addition, the study looked at two types of physicians, general/family practitioners and general surgeons, to determine if the factors differ across types of physicians. These types of physicians were selected because they serve as a major source of referrals for breast cancer patients in Wisconsin, where the study was conducted.

In 1979, Wisconsin had the 16th highest incidence rate of new breast cancer cases per 100,000 population and the 11th highest breast cancer mortality rate per 100,000 population in the nation.[19] There were approximately 1,015 licensed general/family practitioners and 418 licensed general surgeons in office-based practice in the state at the time the study was conducted.[20] These physicians had a number of centers available, located either in Wisconsin or adjacent states and associated with either university medical schools, such as Wisconsin, Minnesota, or Chicago, or large multispecialty group practices, such as the Mayo Clinic and the Marshfield Clinic, for the referral of their patients. In addition, there were a number of oncologists associated with large hospitals in the major cities within or bordering the state which also could be used for patient referral.

Methods

The study consisted of three phases. The first phase identified the factors that appeared to be important in deciding whether or not to refer a patient and where to refer that patient. Given the lack of understanding of the referral process from the literature, two field studies were conducted to gain insight into how, where, when and why patients are referred. In the first study, the author informally interviewed 38 physicians in active practice in Wisconsin, the majority of whom were primary care physicians.[21] The physicians were not selected using any sampling scheme but were physicians who were thought to be open and informative about the referral process and who would be willing to talk to the author about it for an hour. The interviews were content analyzed to identify the factors which the physicians reported as influencing their referral decisions. Because patients often were reported as playing a role in the referral decisions, the author, in the second field study, interviewed 23 patients who had been referred to a major medical center to clarify this role.[22] The results of these studies, as well as the pertinent referral literature, served as the basis for identifying the referral factors addressed in this study.

During the second phase of the study, 11 general surgeons and 14 general practitioners in active practice in Wisconsin rank ordered the relative importance of the factors in making the referral decisions. These physicians either had participated in the physician field study or were nominated by physicians who had participated in the study. Each physician was first given a deck of 3 × 5 cards with one refer/not refer factor and accompanying definition on each card. After reading through all of the cards, the physician rank ordered the factors according to the perceived importance in deciding whether or not to refer a patient. Second, each physician was given a deck of 3 × 5 cards listing the where-to-refer factors and accompanying definitions, and asked to rank order these factors in a similar manner. This task was done on an individual physician basis in the physician's office with no interaction among physicians.

The third phase of the study attempted to predict referral decisions using the identified factors. Two alternative decision making models, which have been proposed and researched extensively as normative models to facilitate the decision making process, were used.[23–25] These models provided alternative ways of looking at a decision-making situation, either through probabilities or utilities, and have been found in some situations to outperform regression models in predicting individual decision-making behavior.[25,31] Also, the use of these models overcome the problem of not having an actuarial data base from which to develop regression models.

The first model was the Multi-Attribute Utility (MAU) model, which in its simple linear form can be expressed as:

$$U_k = W_1U_1(F_1) + W_2U_2(F_2) + \cdots + W_NU_N(F_N)$$

where U_k = the utility or preference associated with the k *th* decision, such as refer or not refer, F_i = the level of the *ith* factor present with the k*th* decision ($i = 1, 2, \cdots, N$), $U_i(F_i)$ = the utility or preference associated with the level of the i *th* factor ($i = 1,2, \cdots ,N$), and W_i = a scaling constant that reflects the relative importance of the i *th* factor on the overall utility of the k*th* decision, such that $0 \leq W_i \leq 1$ and $\Sigma W_i = 1$ ($i = 1,2, \cdots ,N$).

Thus, given the relative importance of each of the referral factors in making a referral decision, such as the decision to refer/not refer, the levels of each of the referral factors present for each alternative decision and the relative preference of each of those factor levels, the MAU model calculates a value representing the relative preference or utility of each alternative decision. The decision with the greatest utility should be the decision that is made. This model has been found to be a valid predictor of decisions regarding health program funding,[26] job search and applicant placement,[27,28] degree of burn severity[29] and physician shortage areas.[30]

The second model was the conditionally independent ratio form of Bayes' Theorem referred to as the Bayesian Probability (BP) model, which can be expressed as:

$$\frac{P(D_1|F_1,F_2, \cdots ,F_N)}{P(D_2|F_1,F_2, \cdots ,F_N)} = \frac{P(F_1|D_1)}{P(F_1|D_2)} \cdot \frac{P(F_2|D_1)}{P(F_2|D_2)} \cdots\cdots \frac{P(F_N|D_1)}{P(F_N|D_2)} \cdot \frac{P(D_1)}{P(D_2)}$$

where

$\dfrac{P(D_1|F_1,F_2, \cdots ,F_N)}{P(D_2|F_1,F_2, \cdots ,F_N)}$ = the posterior odds of decision D_1 being made in comparison to decision D_2 given the levels of factors $F_1, F_2, \cdots ,F_N$ are present, $P(F_i|D_1)$ = the likelihood ratio of the level of factor F_i being present given decision D_1 is made in comparison to decision D_2 ($i = 1,2, \cdots ,N$), $P(D_1)$ = the prior odds of decision D_1 being made in comparison to decision D_2 without the levels of factors $F_1,F_2, \cdots ,F_N$ being present.

Thus, given the prior odds of one alternative decision being made compared with another, such as refer the patient versus not refer the patient, and the likelihood of each of the factor levels being present if one alternative decision is made compared with another, the BP model calculates the odds of one decision being made compared with the other. From this posterior odds, the probability of each of the decisions being made can be calculated by:

$$P(D_1|F_1,F_2, \cdots ,F_N) = \frac{\text{Posterior Odds}}{1 + \text{Posterior Odds}}$$

$$P(D_2|F_1,F_2, \cdots ,F_N) = 1 - P(D_1|F_1,F_2, \cdots ,F_N)$$

The decision with the largest probability is the decision that should be made. Models of this form have been used in the decision making areas of medical diagnosis,[31-33] intelligence gathering,[34,35] job selection[36] and weather forecasting.[37]

Because of the lack of actuarial data regarding the referral factors, the likelihood ratios of the BP model and the factor weights and utilities of the MAU model were subjectively estimated by eight general surgeons (GSs) and eight general practitioners (GPs) in active practice in Wisconsin. These physicians were selected randomly from the physicians who rank ordered the referral factors in the second phase of the study. The remaining physicians were used to pilot test the estimation procedures and, thus, their results were not included in this phase of the study. Eight physicians, four GSs and four GPs, made estimates for the MAU model and eight made estimates for the BP model. Determination of which physicians estimated which parameters was done on a random basis.

For the BP model, each physician was asked to consider a situation where 100 physicians who referred their last breast cancer patient to a consultant were selected randomly. Given this was all the information about these physicians, the patients, and the consultants, the physician estimated how many of these 100 physicians would perceive, for example, that 1) they provide a higher quality of patient management than a consultant, 2) they provide the same quality of patient management as a consultant, and 3) a consultant provides a higher quality of patient management than they. After completing this task, each physician repeated the approach for the situation where 100 physicians did not refer their last breast cancer patient. This task was completed for each of the refer/not refer factors. The likelihood ratio for each factor level was calculated by dividing the likelihood estimate given the decision to refer, by the estimate given the decision not to refer. A similar approach was used to make likelihood estimates for the where-to-refer factors.

For the MAU model, the estimation task was divided into two phases. First, after ranking the relative importance of the refer/not refer factors, each physician rated the relative importance of each factor, assigning the least important factor an arbitrary value of "10" and then giving every other factor a score reflecting its relative importance. For example, if a physician thought a factor was 1½ times as important as the least important factor, a score of "15" would be assigned. No upper bounds were placed on the scores and values reflecting equal importance were permitted. Individual physician factor weights then were calculated by normalizing these ratings. A similar approach was used to estimate the factor weights for the where-to-refer factors.

The second phase required each physician to estimate the utilities of the MAU model by assigning preference scores to each level of the refer/not refer and where-to-refer factors. Each physician was given a preference sheet for each factor, which listed five factor levels. The physician assigned a score of 0 to the least preferred level, a score of 100 to the most preferred level and scores in between reflecting relative preferences for every other level.

Estimates for both models were made on an individual physician basis without interaction. The averages of the estimates across physicians were used in the models.

In addition, 10 GPs and five GSs in active practice in nonmetropolitan areas in Wisconsin each reviewed four hypothetical case descriptions of breast cancer patients, which were based on actual patients. These physicians had not participated in the earlier phases of the study and were selected on the basis of their willingness to provide two hours of time to evaluate the cases. To test the ability of the models to predict referral as well as nonreferral decisions, the cases were designed, with the assistance of a fellow of the University of Wisconsin's Clinical Cancer Center, to be representative of breast cancer patients as they might present themselves in a primary care physician's office, ranging in severity from those that a general practitioner or general surgeon would tend not to refer to those that they would tend to refer immediately. After reading a case description, the physician was asked whether or not he/she would refer the patient and, if so, where. The physician was not provided with a list of alternative referral centers, but indicated his/her first and second choices for consultants. In addition, after reading the definitions of the referral factors, each physician specified the level of each factor involved in making the decisions. This task was completed on an individual physician basis, without interaction, in each physician's practice setting.

The information obtained from the referring physicians on the levels of the factors involved in making the decisions was then used by the models to calculate a predicted decision. This predicted decision was compared with the actual decision made by the referring physician. The per cent of decisions correctly predicted by each of the models was then determined. These percentages were analyzed using a split-plot repeated measures design, with one between-block treatment (type of physician making the referral decisions) and one within-block treatment (type of model).[38] Because the per cent scores were calculated on different sample sizes (five GSs vs. 10 GPS), weighted scores were used in the analyses to insure equal cell variances. A 0.05 significance level was used in all analyses.

Results

The field studies indicated that 12 factors appeared to be important in deciding whether or not to refer a patient and 15 factors appeared to be important in deciding where to refer a patient. The Appendix presents a brief definition of each factor. As illustrated in Tables 1 and 2, several factors appear under both decisions because these factors were perceived by physicians as important issues, regardless of whether the patient was treated by the referring physician or a consultant, or whether the patient was treated by one of several possible consultants. Also, the factors appear to be of three basic types. First are those factors, such as "patient results," which relate to the treatment/care the patient would receive. Second are those noncare related factors, such as "convenience to the patient," which are associated with the patient and/or family. Third are noncare related factors, such as "reciprocations physician receives," which are associated with the referring physician.

TABLE 1. Decision to Refer/Not Refer Factors

Technical/care related factors
Quality of patient management
Patient results
Individualized patient management and care
Patient related factors
Convenience to the patient
Cost to the patient
Patient's potential for productive remaining life
Patient expectations for referral
Physician related factors
Physician's loss of patient/family/income
Personal benefits physician receives
Physician's satisfaction with previous referrals
Referring physician's degree of belief referring is admitting failure
Attitude of physician's colleagues/local medical community towards referring

Table 3 presents the rank orderings of the 11 GSs and the 14 GPs for the decision to refer/not refer factors, along with the average GS and GP rank orderings. There is significant agreement among both the GSs and the GPs as to the relative importance of the factors, with slightly greater agreement among the GPs than the GSs. The Spearman rank correlation coefficient calculated between the average GS and

TABLE 2. Decision Where to Refer Factors

Technical/care related factors
Quality of patient management
Patient results
Individualized patient management and care
Patient related factors
Convenience to the patient
Cost to the patient
Degree patient preferences are met
Previous use of and satisfaction with consultant by patient/family
Physician related factors
Physician's loss of patient/family/income
Reciprocations physician receives
Communication physician receives
Respect/courtesy physician receives
Attitude of patients towards consultant
Physician's personal knowledge of consultant
Physician's satisfaction with previous referrals to consultant
Attitude of physician's colleagues/local medical community towards consultant

average GP rank orderings indicates a significantly high level of agreement between the two types of physicians. The most important factors for both types of physicians were the technical/care related factors of "quality of patient management," "patient results" and "individualized patient management and care," followed by "patient's potential for a productive remaining life." The least important factors for both physicians were the physician-related factors of "referring physician's degree of belief referring is admitting failure," "physician's loss of patient/family/income" and "attitude of physician's colleagues/local medical community towards referring."

Similarly, Table 4 presents the rank orderings for the decision where-to-refer factors. Inspection of Table 4 indicates again significant agreement both among GSs and GPs as well as between GSs and GPs as to the relative importance of the factors. Again, those factors related to the treatment aspects of a patient's care were ranked most important as well as the "physician's personal knowledge of the consultant." The least important factors were the physician-related factors of "physician's loss of patient/family/income," "attitude of physician's colleagues/local medical community towards consultant," "cost to the patient," "reciprocation physician receives," and "respect/courtesy physician receives."

Table 5 indicates that, on the average, the MAU and BP models using all of the referral factors were able to correctly predict 74.2 per cent of the decisions to refer/not refer and 81.0 per cent of the decisions of where to refer. The analyses of variance revealed highly significant differences between the models for the decisions to refer/not refer, with the BP model correctly predicting a significantly greater per cent of the GS decisions than the GP decisions (90.0 per cent vs. 70.0 per cent), while the MAU model correctly predicted a significantly greater per cent of the GP decisions than the GS decisions (77.5 per cent vs. 60.0 per cent). For the decision where to refer, no significant differences between models were found. On the average, the models correctly predicted a greater per cent of GS decisions than GP decisions (91.7 per cent vs. 76.7 per cent).

However, models using fewer referral factors were found to predict better than or as well as the models using all of the referral factors (Table 6). For the decision to refer/not refer, a three factor MAU model (quality of patient management, patient results, and attitude of physician's colleagues/local medical community towards referring) and a four factor BP model (quality of patient management, patient expectations for referral, referring physician's degree of belief referring is admitting failure and attitude of physician's colleagues/local medical community towards referring) were found to predict a greater per cent of the decisions correctly than the respective total factor models. No significant differences were found between these fewer factor models nor were differences found in the performances of the models between GS and GP decisions.

What is of interest are the factors included within these models. Both models included the factor "quality of patient management," which was ranked as the most important factor by both the GSs and GPs and the factor "attitude of physician's colleagues/local medical community towards referring," which was ranked third least important by the GS and GP physicians. In addition, the BP model included the factor "referring physician's degree of belief referring is admitting failure," which was ranked least important by the physicians.

To further address this issue, a stepwise regression analysis was performed to identify the set of factors that best explained the decisions, using the ratings of the factors provided by the referring physicians as the independent variables and the referring physicians' decisions (refer vs. not refer) as the dependent variables. "Attitude of

TABLE 3. General Surgeon (GS) and General Practitioner (GP) Rankings of the Decision to Refer/Not Refer Factors (12—Most Important, 1—Least Important)

Factors	General Surgeons												General Practitioners														
	1	2	3	4	5	6	7	8	9	10	11	Average	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Average
Quality of patient management	12	9	12	8	12	6	11	12	10	12	12	12	12	11	12	10	11	10	11	12	12	12	12	10	12	10	12
Patient results	11	11	6	5	10	11	8	9	9	11	11	11	4	12	9	12	12	9	10	9	11	10	2	9	11	8	10
Individualized patient management and care	9	10	3	7	9	5	12	11	11	10	10	10	11	10	10	7	10	11	8	11	10	11	11	7	10	9	11
Convenience to the patient	6	5	11	11	8	3	6	5	8	8	7	8	9	6	6	3	6	8	6	10	9	5	7	5	6	6	6
Cost to the patient	5	6	5	10	7	9	4	4	7	7	6	5	5	3	7	5	7	7	7	8	8	4	5	6	7	5	5
Physician's loss of patient/family income	1	1	1	1	2	4	3	3	1	6	1	2	3	2	2	4	3	1	1	1	2	1	3	3	3	2	2
Personal benefits physician receives	7	4	9	4	5	7	7	7	3	5	5	4	8	7	5	6	2	2	5	4	4	8	6	4	4	4	4
Patient's potential for productive remaining life	10	12	4	3	11	8	5	8	12	9	9	9	10	8	11	11	8	3	12	6	7	9	9	11	8	12	9
Patient expectations for referral	4	3	7	12	4	12	10	6	6	1	8	6	7	9	8	8	5	12	4	7	6	6	8	12	9	7	8
Physician's satisfaction with previous referrals	8	8	10	6	6	10	9	10	4	2	4	7	6	5	3	9	9	6	9	5	5	7	10	8	5	11	7
Referring physician's degree of belief referring is admitting failure	3	2	2	2	1	2	2	2	2	3	2	1	1	1	1	1	1	4	2	2	1	2	1	1	1	3	1
Attitude of physician's colleagues/local medical community towards referring	2	7	8	9	3	1	1	1	5	4	3	3	2	4	4	2	4	5	3	3	3	3	4	2	2	1	3
Kendall Coefficient of concordance												0.5464 (p<.001)															0.7328 (p<.001)

Spearman Rank Correlation Coefficient between average GS and GP rankings = 0.9650 (p < .01).

TABLE 4. Continued.

Physician's personal knowledge of consultant	12	11	12	8	11	12	13	12	12	12	6	13	11	13	12	11	11	15	12	9	8	12	6	7	12	7	12
Physician's satisfaction with previous referrals to consultant	8	12	14	9	10	11	10	11	11	11	7	11	10	9	6	9	12	13	11	5	7	13	7	6	14	11	10
Attitude of physician's colleagues/local medical community towards consultant	2	3	8	7	8	1	2	1	5	1	4	2	1	2	4	1	8	3	2	3	3	2	1	2	2	4	2
Kendall Coefficient of Concordance												0.5821 (p<.001)															0.6368 (p<.001)

Spearman Rank Correlation Coefficient between average GS and GP rankings = 0.9759 ($p < .01$).

TABLE 5. Mean Per Cent Correct for the Total Factor MAU and BP Models by Type of Referring Physician for the Decisions to Refer/Not Refer and Where to Refer

Decision	Referring Physician	MAU Model	BP Model	Average
Refer/not refer	GS	60.0	90.0	75.0
	GP	77.5	70.0	73.8
	Total	71.7	76.7	74.2
Where to refer	GS	91.7	91.7	91.7
	GP	76.7	76.7	76.7
	Total	81.0	81.0	81.0

physician's colleagues/local medical community towards referring" was found to be the best explainer variable, followed by "patient expectations for referral" and "quality of patient management." These three factors explained approximately 41 per cent of the variance in the decisions. Two of these factors "attitude of physician's colleagues/local medical community towards referring" and "quality of patient management" were included in the three factor MAU model and all three factors were included in the four-factor BP model.

Similarily, for the decision of where to refer, a five-factor MAU model (degree patient's preferences are met, reciprocations physician receives, communication physician receives, previous use of and satisfaction with consultant by patient/family, and physician's personal knowledge of consultant) and a four-factor BP model (same factors as the five factor MAU model, except communication physician receives) were found to correctly predict a greater per cent of the decisions than the respective total factor models (Table 6). No significant differences were found between the fewer factor models nor were differences found in the performances of the models between GS and GP decisions.

Interestingly, the technical/care-related factors were absent from the fewer factor models for the decision of where to refer. An explanation for this may be related to

TABLE 6. Mean Per Cent Correct for the Fewer Factor MAU and BP Models by Type of Referring Physician for the Decisions to Refer/Not Refer and Where to Refer

Decision	Referring Physician	MAU Model	BP Model	Average
Refer/not refer	GS	85.0	85.0	85.0
	GP	85.0	90.0	87.5
	Total	85.0	88.3	86.6
Where to refer	GS	91.7	83.3	87.5
	GP	80.0	83.3	81.6
	Total	87.5	83.3	85.4

one of the findings of the physician field study. During that study, physicians indicated that the treatment/care-related factors were the most important factors and would be considered first in making the referral decisions. If those factors were equivalent for various consultants, then the other factors would play a major role in the decisions. If this is the case, and the consultants, who were specified by the physicians when they indicated where they would refer patients, were perceived as equivalent in terms of the technical/care-related factors, then these factors should be absent from the models. Lack of variability of the dependent variable, type of decision, prohibited a stepwise regression analysis for the decision where to refer.

Conclusions

The results of the study have indicated that there are a number of factors, most of which have not been studied in the referral literature, that appear to be important when a physician is faced with the possibility of referring a patient. Of prime importance are factors related to the kind and quality of medical care the patient would receive. Although there seem to be some individual differences, generally there appears to be a high level of agreement between and among different types of referring physicians as to the relative importance of the factors. Also, there did not seem to be any additional support for Shortell and Vahovich's hypothesis that patient-related factors may be more important to client-dependent physicians, such as general practitioners, and physician-related factors may be more important to colleague-dependent specialists, such as surgeons.

The normative decision-making models correctly predicted a relatively high per cent of referral decisions. If the models, whose parameters were subjectively estimated, have some internal validity, these results lend some credibility to the identified factors. However, it also was found that not all of the factors appear to be important in predicting the referral decisions. The factors found to be the best predictors, through the models and a stepwise regression analysis, were not always consistent with the factors reported by the physicians to be the most important factors. These results raise questions about differences between what physicians may report to be important, what they may actually think is important and how they behave, which must be investigated further.

It is interesting to note the apparent importance of the factor "attitude of physician's colleagues/local medical community towards referring" in predicting the decisions to refer/not refer. As suggested by physicians interviewed during the field studies, a physician's colleagues may have a significant impact on his/her referral behavior to the point of determining which patients will be referred. Because of the implications such norms may have, this issue needs further investigation.

The results of this study suggest major implications for consulting physicians or centers. Programs for accepting patient referrals, diagnosing and/or treating those patients and returning the patients to the referring physicians, should be structured and operated in such a way that recognize the factors that a consultant's potential clientele consider important in the referral process. The results of this study imply that it is imperative for a consultant to first have the technical capability and expertise to provide high quality patient management with positive patient results, in order to maintain a viable referral network. However, the consultant must then address the other physician- and patient-related factors of importance to the referring physicians. Consultants should attempt to personalize their relations with referring physicians. The results of the study indicate that personal knowledge of the consultant appears to be an important consideration when referring physicians decide where to refer patients. This implies that consultants should take some initiatives in overcoming the town-and-gown syndrome by, for example, visiting potential referring physician clientele and helping those physicians better address their problems. Some referring physicians expect consultants to reciprocate patient referrals through education on the diagnoses and/or treatment of various types of cases, training of support staff personnel and other ways of improving his/her capabilities. The study also indicated that the referring physician's colleagues and/or local medical community play a role in the referral process, implying the need for consultants to educate the medical community regarding the diagnosis and treatment of various types of problems, the appropriate use of referrals and their capabilities and expertise. This could be accomplished, for example, through continuing education programs jointly sponsored by the consultants and local medical societies.

The results of the study also indicate that the patient does play a role in the referral process. Referring physicians appear to be attuned to the patient's expectations for referral, preferences for certain consultants or referral centers, and previous use of and satisfaction with consulting physicians or centers. This implies that consultants must be aware of this role, realizing that the patient provides an important feedback mechanism not only back to the referring physician but also to friends and neighbors. The consultants need to insure that services are provided to patients in such a way that every patient is a potential "sales representative" of their services. In addition, the study implies that consultants should educate the patient in appropriate referral behavior to insure that the patient does not place inappropriate referral demands on referring physicians.

This study found that little actually is known about the referral process in terms of how it operates; the relationships between referring physicians, patients and consultants; the decision making processes of each of the key participants and the environmental context which influences the operation of the process. Research is needed to understand better the transactions between the referring physician and the patient regarding the referral decisions, and the transactions between the referring and consulting physicians during a referral encounter. Although this study dealt with the referral of breast cancer patients, the applicability of the results and conclusions may extend beyond this limited sample of referrals to the general population of patient referrals. The results of the field studies indicated that the referral factors addressed in this study are applicable to any type of patient referral.[21] However, what is not known is the role these factors play in different types of referral decisions. For example, the relative importance of the factors may change for different types of referral patients. Further research must be conducted to further validate the findings of this study in order that the results may be generalized to other types of patients and referral transactions. Also, research is required on developing measures of effectiveness, not only in terms of process but also outcome, that can be used to assess and evaluate the referral process and to make appropriate modifications to maximize a referral program's efficiency and effectiveness.

References

1. Metcalfe DHH, Sischy D. Patterns of referral from family practice. J Fam Prac 1974;1:34-8.

2. Penchansky R, Fox D. Frequency of referral and

patient characteristics in group practice. Med Care, 1970;8:368-84.

3. Shortell SM. A model of physician referral behavior: a test of exchange theory in medical practice. Research series no 31. Chicago: Center for Health Administration Studies, University of Chicago, 1972.

4. Avnet HH. Physician service patterns and illness rates: a research report on medical data retrieved from insurance records. New York: Group Health Insurance, Inc, 1967.

5. Geyman JP, Brown TC, Rivers K. Referrals in family practice: A comparative study by geographic region and practice setting. J Fam Pract 1976;3:163-7.

6. Ruane TJ. Consultation and referral in a Vermont family practice: A study of utilization, specialty distribution, and outcome. J Fam Pract 1979;8:1037-40.

7. Schmidt DD. Referral patterns in an individual family practice. J Fam Pract 1977;5:401-3.

8. Shortell SM, Vahovich SG. Patient referral differences among specialities. Health Serv Res 1975;10:146-61.

9. Taubenhaus LJ. A study of one rural practice, 1953. General Practitioner 1955;12:96-102.

10. Tenney JB, White KL, Williamson JW. National ambulatory medical care survey: background and methodology, (Vital and Health Statistics, Series 2, No 61) DHEW Publication: HRA 74-1335, 1974.

11. Williams TF, White KL, Fleming WL, Greenberg BG. The referral process in medical care and the university clinic's role. J Med Educ 1961;36:899-907.

12. Clute KF. The general practitioner: a study of medical education and practice in Ontario and Nova Scotia. Toronto: University of Toronto Press, 1963.

13. Miller MH. Who receives optimal medical care? J Health Soc Behav 1973;14:176-82.

14. Peterson OL, Andrews LP, Spain RS, Greenberg BG. An analytical study of North Carolina general practice. J Med Educ 1956;31:1-165.

15. Bates BB, Torkelson LO. Two methods of improving referral information: a comparative study in a university medical center. Med Care 1967;5:418-22.

16. Cummins RO, Smith RW. Reports to referring physicians. N Engl J Med 1975;293:361, 362.

17. Kunkle EC. Communication breakdown in referral of the patient. JAMA 1964;187:663.

18. Williams TF, White KL, Andrews LP, Diamond E, Greenberg BG, Hamrick AA, et al. Patient referral to a university clinic: Patterns in a rural state. Am J Public Health 1960;50:1493-1507.

19. American Cancer Society. Cancer facts and figures, 1979. New York: American Cancer Society, 1979.

20. American Medical Association. Physician distribution and medical licensure in the US. Chicago: American Medical Association, 1977, 1979.

21. Ludke RL. Physician referrals: results of a field study of physicians. Madison: University of Wisconsin-Madison, 1977, unpublished report.

22. Ludke RL. Physician referrals: results of a field study of patients. Madison: University of Wisconsin-Madison, 1977, unpublished report.

23. Edwards W. Behavioral decision theory. Ann Rev Psychol 1962;12:473-98.

24. Slovic P, Fischoff B, Lichtenstein S. Behavioral decision theory. Ann Rev Psychol 1977;28:1-39.

25. Slovic P, Lichtenstein S. Comparison of Bayesian and regression approaches to the study of information processing judgment. Organizational Behavior and Human Performance 1971;6:649-744.

26. Stimson D. Utility measurement in public health decision making. Management Science 1969;15:17-30.

27. Fischer G. Convergent validation of decomposed multi-attribute utility assessment procedures for risky and riskless decisions. Organizational Behavior and Human Performance 1977;18:295-315.

28. Huber GP, Daneshgar R, Ford D. An empirical comparison of five utility models for prediciting job preferences. Organizational Behavior and Human Performance 1971;6:267-82.

29. Gustafson DH and Holloway D. A decision theory approach to measuring severity in illness. Health Serv Res 1975;10:97-106.

30. Health Services Research Group, Center for Health Systems Research and Analysis, University of Wisconsin. Development of the index of medical underservice. Health Serv Res 1975;10:168-80.

31. Gustafson DH. Evaluation of probabilistic information processing in medical decision making. Organizational Behavior and Human Performance 1969;4:20-34.

32. Gustafson DH, Kestly J, Ludke RL, Larson F. Probabilistic information processing: implementation and evaluation of a semi-PIP diagnostic system. Comput Biomed Res 1973;6:355-70.

33. Greist J, Gustafson DH, Stauss F, Rowse G, Laughren T, Chiles J. A computer interview for suicide-risk prediction. Am J Psychiatry 1973;130:1327-332.

34. Edwards W, Phillips L. Man as transducer for probabilities in Bayesian command and control systems. In: Bryon GL, Shelly MW, eds. Human judgments and optimality, New York: John Wiley and Sons, 1964.

35. Kelly C, Peterson C. Probability estimates and probabilistic procedures in current-intelligence analysis. IBM Rep 71-5047. Gaithersburg, Md.: IBM 1971.

36. Ford D, Huber GP, Gustafson DH. Predicting job choices with models that contain subjective probability judgments: an empirical comparison of five models. Organizational Behavior and Human Performance 1972;17:397-416.

37. Julian P, Murphy A. Probability and statistics in meteorology: a review of some recent developments. Bulletin of the American Meteorological Society 1972;53:957-63.

38. Kirk R. Experimental design: procedures for the behavioral sciences. Belmont, Calif.: Wadsworth, 1968.

Appendix

Definitions of the Referral Factors

Technical/Care-Related Factors

Quality of Patient Management. Thoroughness of workup and accuracy of staging; thoroughness of problem/treatment explanation to patient/family; thoroughness and efficiency of treatment; comprehensiveness of rehabilitation.

Patient Results. How patients look; complication rate; pre, during and post treatment mortality rates; cure rates.

Individualized Patient Management and Care. Management and care of the patient provided according to the specific needs and condition of the patient in a friendly, empathetic, personal manner.

Patient-Related Factors

Cost to the Patient. Travel expenses incurred by the patient/family in travelling to receive care including food and lodging; other expenses involved, e.g., babysitting costs, lost wages, etc.; out-of-pocket expenses for care.

Convenience to the Patient. Distance; degree of physical discomfort involved in travelling; ease in locating where to go and parking; availability of suitable accommodations nearby; extent of disruption of patient's/family's home life.

Patient's Potential for Productive Remaining Life. Patient's life expectancy and quality of remaining life.

Patient Expectations for Referral. Patient's and/or family's implicit or explicit desire to be referred or not referred.

Degree Patient's Preferences are Met. Expressed patient preferences regarding consultant; convenience such as distance or ability to stay with family or friends, costs.

Previous Use of and Satisfaction with Consultant by Patient/Family. Patient/family having used consultant for same, related, or nonrelated problems and their level of satisfaction with consultant.

Physician-Related Factors

Physician's Loss of Patient/Family Income. Patient/Family not returning to physician for care; consultant performing procedures which physician could perform.

Personal Benefits Physician Receives. Help with problem; education about problem diagnosis and treatment; reciprocations; avoiding hassles of referring.

Physician's Satisfaction with Previous Referrals. Satisfaction with results; the way patients were treated; the way the physican was treated.

Referring Physician's Degree of Belief Referring is Admitting Failure. Extent to which the physician believes that referring a patient is openly admitting he/she is no longer able to care for that patient.

Attitude of Physician's Colleagues/Local Medical Community Towards Referring. Norms that the physician's colleagues and local medical community have established for referring or not referring type of patient.

Reciprocations Physician Receives. What the physician may gain by referring patients to the consultant such as education, consultant coming to hold special clinics to train personnel in special procedures, patient referrals, etc.

Communication Physician Receives. Detail and quality of the reports; the promptness of the reports; admission and discharge notices; and notices of crises that arise.

Respect/Courtesy Physician Receives. Consultant's attitude towards helping and towards the physician in front of patients and others.

Attitude of Patients Toward Consultant. How satisfied/dissatisfied physician's patients previously referred to consultant were with that consultant.

Physician's Personal Knowledge of Consultant. Having worked or trained with consultant or know personally through some other means such as previous referrals, meetings, conferences, etc.

Attitude of Physician's Colleagues/Local Medical Community Towards Consultant. Norms that the physician's colleagues and local medical community have established for using that consultant for the patient.

Physician's Satisfaction with Previous Referrals to Consultant. Satisfaction with results; the way patients were treated; the way the physician was treated.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Use of Illness-Related Ambulatory Physician Services in Finland

PERTTI KEKKI, M.D., SC.D., DCM(EDIN)

The total utilization rate of ambulatory illness-related physician services in Finland is higher in big urban health center areas than in rural areas. When only the health center physician utilization rates were studied the finding was opposite. A notable finding related to the impact of the national health planning were the surprisingly equal average utilization rates of public sector health services (health center physician and general hospital outpatient services combined) in both areas. In fact, use of cheaper public services was greater in rural than in urban areas. The difference in total use rates was almost entirely due to the larger use of more expensive private physician services in urban areas. At present, there is a great pressure for increasing the supply of health center services in urban areas to the same level as in rural districts. One of the fundamental questions seems to be the pricing of the two service sectors—public and private.

IN ALMOST ALL of the industrialized countries, efforts have been made to improve the availability and accessibility of health services in order to be able to provide the necessary services to the population, regardless of the place of residence or income status.

The common finding, however, has been that even in countries where the above goals have been stated in the legislation, there are continous inequalities. In England, for example, it was found that the inequalities between the regions had increased rather than decreased during the first 21 years of the N.H.S. Those regions which previously had many services available, had gained more, or the differences had remained the same.[1] Hart[2] presented the Inverse Care Law, according to which, in England, the availability of the high quality physician services was inversely related to the need of the population for those services.

In Finland, the goals of the universal sickness insurance (in effect from 1964)[3] and the Community Health Care Act of 1972, which emphasized the importance of primary care and created the health center network covering the whole country,[4,5] were to make the population groups living in various parts of the country equal as for the possibilities to use health services.

The development of health services after 1972 is based on five-year health planning, according to which the public sector health resources are being allocated. In this planning mechanism, there is an in-built evaluation system. However, despite the various planning activities, the factual distribution of the services or their use have not been widely examined. This is due to the gap in the planning system. To get a real picture

---

Institute of Public Health Science, University of Tamperg and Department of Primary Medical Care, University of Helsinki.

Supported in part by grants from the Finnish government and the Yrjö Jahnsson Foundation (Finland).

The basic work for this study was carried out in collaboration with the Research Institute for Social Security of the Social Insurance Institution of Finland at the Division of Health Care Organization of the Johns Hopkins University School of Hygiene and Public Health in Baltimore, Maryland.

Reprint requests: Pertti Kekki, M.D., Sc.D., DCM(Edin), Assistant Professor (Health Services Administration) University of Tamperg and Acting Professor of Primary Medical Care, University of Helsinki. Department of Primary Medical Care, University of Helsinki, 2. Linja 4C, SF-00530 Helsinki 53, Finland.

0025-7079/82/0800/0797/$01.40 © J. B. Lippincott Co.

of the developments and utilization of the various services at the primary level is impossible, since the units examined are large aggregates (provinces or central hospital districts). At the provincial level there are, of course, currently better possibilities for studying smaller units, such as the individual health center areas. The difficulty there lies in the data collected routinely because these do not allow for in-depth analyses as some more specific data are not routinely collected.

In some studies (e.g., Tuomikoski[6], it has been suggested that planning has not brought about the solution in the distribution or clustering of the services.

Apart form the public sector services (health centers and the great majority of hospitals), people in Finland can also use the services of the private sector. The basic difference between the health center and private sector illness-related physician services is that the physicians working in the former are salaried medical officers and the physicians working in the latter are usually hospital physicians who when working in the hospitals are full-time salaried staff but, who after their regular working hours, turn to private practitioners working on a fee-for-service basis in private medical centers or group practices outside the hospitals. The majority of the private practioners are specialists whereas the health center physicians are usually general practitioners. In 1975, about 60% of all physicians working in hospitals were in private practice, the proportion being probably greater for physicians working in general hospitals. Only about 10% of all physicians in the country were full-time private practitioners.

Sickness Insurance reimbursed in 1975 about 44% of the private physician fee to the patient afterwards, the average fee being about 31 FIM(US$ 7.50). In health centers, the patient paid about US$0.80 for a health center physician visit in 1975. The differences will be even greater if the fees for diagnostic services (lab, x-ray) are taken into account. From the beginning of 1976, laboratory and x-ray services in health centers have been free to the patient, and since the beginning of 1981 the physician visit also is "free."

Goals

The general goal of the present study is to increase the knowledge and understanding of health services in Finland. For planning purposes, it is considered useful to see whether a mechanism of national health planning, acting in a well-defined health services system in a developed country, has been able to increase equity with regard to use of physician services in ambulatory care.

The specific aims of this research were: 1) to study whether, in 1975, there still were considerable differences between urban and rural areas with regard to the volume of use of ambulatory physician services; 2) in case differences that exist, to study also the variation of use within the groups of urban and rural areas and the possible reasons for these differences; and 3) to correlate the findings with the goals of the national health policy and health planning.

It is especially important to note that, in this study, we use as the units the basic units of the Finnish health services—the health center areas, in contrast to other studies which have used the basic administrative units—the communes or local authority areas. Most of the health center areas are composed of more than one commune. This fact makes the findings more reliable from the health planning point of view.

Materials and Methods

The material for this study comes from a data set which has been described in detail elsewhere.[7] The material was processed so that the final data set included all of the basic units of the Finnish health services, the health center areas existing in 1975 (216 in total). From each health center area, a large number of resource, demographic, social structure, utilization, and other variables were included. The data, thus in each case, referred to the mid-year population of the particular health center area. The original data were precollected and came from various sources such as the Social Insurance Institution, the Finnish Hospital Association, the Central Statistical Office, the Office of the Cabinet, and the Finnish Medical Association.

In 1975, there were ten health center areas in Finland with a population of 60,000 or more. Conversely, 72 per cent of all the health center areas had a mid-year population of 20,000 or less, and 85 per cent of all 30,000 of less. This study includes 38 health center areas, of which 19 represent the big population centers. These include eight of the areas with a population of 60,000 or more. All of these big centers except one (Vantaa) have a central hospital located in the area (five of them have a university teaching hospital). The other 19 centers are mainly rural health center areas with a population 20,000 or less. This paper examines the use of ambulatory illness-related physician services by the populations of these health center areas. Figure 1 shows the distribution of the centers of the areas included in this study.

Results

Table 1† shows the coverage of illness-related physician services for the health center physicians and private physician services. This is done by presenting the proportions of the populations in each health center area that did not use these services in 1975 (the proportion of the nonusers).

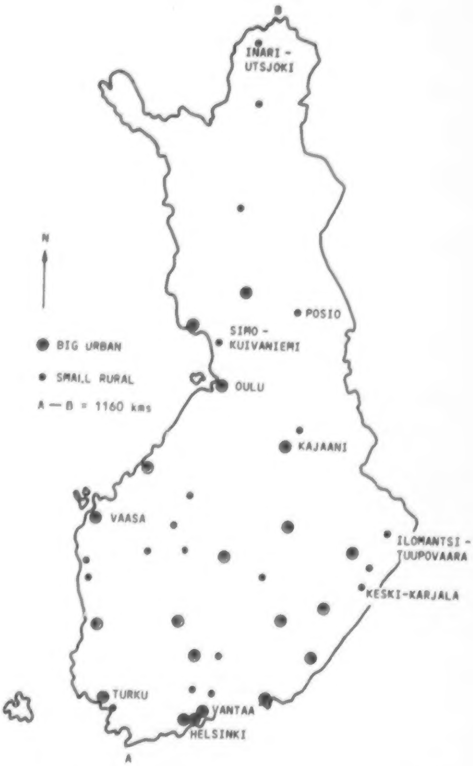


FIG. 1. Map of Finland showing the distribution of the centers of the health center areas included in the study.

The differences between the health center areas were considerable. Whereas 31 per cent of the population in Helsinki (the capital city) used health center physician services in 1975, 60 per cent of the population of the rural health center area of Keski-Karjala (in Eastern Finland) used these services. Conversely, about 43 per cent of the population in Helsinki used private physician services, whereas in the health center area of Keski-Karjala, only 14 per cent of the population received these services. It is, however, interesting to notice that the total percentages of users in both places are equal (74 per cent). In

---

† In these tables the coefficient of variation (C%) also has been calculated. C% = 100 × SD/X. This coefficient is measure of relative variability whereas the standard deviation is a measure of absolute variability.

TABLE 1. Percentages of the Nonusers of Services in the Population by Health Center Area

Large Health Centers	HC-Physician	Private Physician	Small Health Centers	HC-Physician	Private Physician
Helsinki	68.8	57.3	Alajärvi	51.8	72.0
Oulu	74.5	59.9	Jalasjärvi	41.2	88.7
Turku	83.3	49.5	Närpiö	49.6	87.0
Tampere	75.2	58.6	Perhonjokilaakso	51.9	83.3
Kuopio	66.6	55.9	Länsi-Turunmaa	54.6	68.2
Jyväskylä	66.3	62.1	Kiihtelys-Pyhäselkä	48.6	87.3
Joensuu	55.4	74.4	Keski-Kainuu	48.0	84.2
Pori	72.6	53.1	Ilomantsi-Tuupovaara	44.8	92.5
Savonlinnanseutu	53.6	74.6	Keski-Karjala	39.3	85.7
Mikkelinseutu	56.1	72.2	Simo-Kuivaniemi	65.6	80.4
Kotka	61.9	62.2	Sodankylä	39.4	91.4
Vaasa	77.4	65.0	Posio	40.9	87.8
Rovaniemi	58.1	69.4	Ähtärinjärvi	59.0	73.7
Lappeenranta	57.4	69.6	Kristiinanseutu	71.5	70.1
Hämeenlinnanseutu	63.3	62.6	Tuusula	46.0	75.2
Kajaani	54.5	72.1	Inari-Utsjoki	42.0	90.7
Kokkolanseutu	66.2	66.6	Nurmijärvi	44.0	73.8
Vantaa	59.8	58.3	Lamminseutu	44.3	83.3
Kemi	58.5	74.9	Kangasniemi	44.0	84.4
$\overline{X}$	64.7	64.1	$\overline{X}$	48.8	82.1
SD	8.7	7.7	SD	8.8	7.6
C (%)*	13.4	12.0	C (%)	18.0	9.3

* C (%) = coefficient of variation. $C = 100\left(\frac{SD}{\overline{X}}\right)$.

Ilomantsi-Tuupovaara health center area, another rural area situated in North Karelia, only 7.5 per cent of the population had contacted a private physician, and in Lapland (in the Northern part of the country) nine per cent of the population of the Inari-Utsjoki rural health center area contacted a private physician in 1975. When the mathematic means of the utilization rates for the health center and private physicians are calculated, they show that in the small health center group roughly about 50 per cent of the population used health center physician services, compared with the average of about 35 per cent in the group of big centers. As for private physicians, roughly 19 per cent of the populations of the small centers used these services, compared with the about 36 per cent in the big centers. These figures compared well with the averages for the entire country: in 1975, about 43 per cent of the total population used health center physician services—about 37 per cent of the population living in urban areas and 51 per cent of the population living in rural areas. About 30 per cent of the total population used private physician services, about 37 per cent of the population in urban areas and 20 per cent in rural areas. About 41% of the total population did not use either of these two services in 1975. This indicates that some overlapping takes place. Some people have used both the HC and private physician services. In the entire country, about 29 per cent of the population used only health center physician services in 1975, 16 per cent used only private physician services and 14 per cent used both. Over half of the visits made to private physicians were made to specialists. Table 2 presents the utilization rates by health center area for health center physician visits, private physician visits, and outpatient

TABLE 4. General Surgeon (GS) and General Practitioner (GP) Rankings of the Decision Where to Refer Factors (15—Most Important, 1—Least Important)

Factors	General Surgeons 1	2	3	4	5	6	7	8	9	10	11	Average	General Practitioners 1	2	3	4	5	6	7	8	9	10	11	12	13	14	Average
Quality of patient management	15	15	15	11	15	3	15	15	14	15	15	15	15	14	15	14	14	12	14	15	15	15	15	13	15	14	15
Patient results	14	14	13	4	14	7	14	13	13	14	14	14	14	15	13	15	15	4	15	11	14	10	8	14	13	12	13.5
Individualized patient management and care	13	13	6	10	13	5	11	8	15	13	13	12	13	12	14	13	13	11	13	14	13	14	14	8	8	13	13.5
Convenience to the patient	4	7	5	13	5	8	9	4	9	9	11	7	6	4	8	2	5	8	4	12	5	4	13	9	3	8	6
Cost to the patient	3	4	2	3	4	2	3	3	8	6	10	3	5	3	5	3	7	9	3	10	1	3	11	10	11	5	5
Degree patient's preferences are met	9	6	3	12	12	15	12	7	10	10	8	10	9	10	9	12	6	5	7	13	12	5	12	11	9	15	11
Physician's loss of patient/family/income	1	1	1	1	1	4	1	2	1	5	1	1	3	1	2	5	1	1	1	1	2	1	3	1	1	1	1
Reciprocations physician receives	6	2	10	2	3	6	5	6	2	4	3	4	2	7	3	4	2	7	6	2	4	6	9	5	7	2	3
Communication physician receives	5	5	7	5	9	10	8	9	6	3	9	6	7	5	7	10	4	10	5	8	11	9	10	4	10	9	7
Respect/courtesy physician receives	10	10	4	6	2	9	7	5	7	2	2	5	12	6	1	8	3	6	9	4	10	7	2	3	4	3	4
Previous use of and satisfaction with consultant by patient/family	7	8	11	14	6	14	4	14	3	7	12	9	4	8	10	7	9	14	10	6	9	11	4	12	6	6	9
Attitudes of patients toward consultant	11	9	9	15	7	13	6	10	4	8	5	8	8	11	11	6	10	2	8	7	6	8	5	15	5	10	8

visits to general hospitals, as well as the total ambulatory physician utilization rates.

The health center physician visit rate varies from 361/1000 to 2122/1000 in Posio (a rural health center area in the province of Lapland). Knowing also the population bases of the areas the findings suggest that the smaller the central hospital city the higher the health center physician visit rate seems to be (the smaller the private sector and probably the more developed the health center sector). The weighted average for the health center physician visit rates in the areas with more than 60,000 inhabitants is 643/1000, and for the first ten of the smaller health centers in the list this figure is 1440/1000. The weighted average for the private physician visit rates in the large center group is 1389/1000, compared with the 344/1000 in the small center group. In the big center group, the average visit rate to the outpatient departments of general hospitals was 929/1000 while this rate for the small centers is 329/1000. When all these rates are added up for the 19 large and small centers, the resulting total population weighted average use rates are 2819/1000 in the large centers and 1931/1000 in small centers. The difference is almost one visit per person in favor of the big population centers. The emphasis in the big centers is strongly on the private physician sector and, on the other hand, on use of the outpatient services of the big hospitals in these areas. In these areas, the population, as the consequence, uses considerably more specialist services than in the small health center areas (probably for primary care?). The highest total use rate among the health center areas listed here was in Turku (the third largest city in the country on the Southwestern coast) 3.3 visits/person. The next highest was in Helsinki, 3.2 visits/person. The lowest the figure was in the rural health center area of Simo-Kuivaniemi in the sparsely populated Lapland, 1.7 visits/person.

Table 3 presents the utilization rates per 1000 people in three age groups for health center and private physician services. (Hospital outpatient visit rates were not available by age groups.) This table shows interesting differences between the two groups, but also within the groups. For example, in the large centers, the health center physician visit rate for children varies from 855 in Vantaa to 360 in Oulu (a university city in the North). In the small health center group, the range is from 1227 in the area of Inari-Utsjoki to 355 in Simo-Kuivaniemi, both of these areas in Lapland. The rate of private physician contacts varies in the large center group from 829 in Turku to 116 in Kajaani (a relatively small central hospital city in the North) and in the small health center group from 302 to 23. Similar differences can be found in the other age groups.

When the total age-specific utilization rates (health center and private physician use combined) are calculated the variation seems to be smaller in the group of large centers. The highest combined average contact rates are found in the group of the eight large centers having a population of 60,000 or more. In this group, the city of Vantaa has the highest health center physician visit rate for children of all the 38 health centers under study. (Vantaa is a big neighboring city of Helsinki.) Also, in other age groups, the health center visit rates in Vantaa rank high. In the age group of 16–64 years, Vantaa is third after Posio, in Lapland, and Nurmijärvi, near Helsinki. In the age group for the elderly Vantaa is second only to Posio (Posio 3.9 visits/person versus 3.6 visits/person in Vanaa).

In the central hospital cities, the smallest use rates are found in the city of Vaasa (on the West coast about 400 km from Helsinki), where the total utilization rates in each age group are the lowest in the group. The reason for this seems to be the low health center physician use rates, which the private sector is not able to compensate. Of the small health centers, the worst situation seems to be in Simo-Kuivaniemi, where the utilization figures for each age group are lowest.

The most interesting health center of those studied seems to be the health center area of Posio, where the utilization rates for the working age population and the elderly are the highest among the 38 centers, and the utilization rate for the children is also at a relatively high level. The high use rates in Posio are mainly due to the high health center physician visit rates (Table 3). Since in Posio especially the rates for the working age population and the elderly are in their own class, the next step in this study was to examine more closely the health center physician visits in order to find perhaps some clues about the possible reasons for the big differences.

A high use rate may be a consequence of a high coverage of services i.e., the high proportion of the population who has used the services or it may also be a consequence of a high visiting rate of the same persons (frequent visitors). According to Table 1, the coverage of the health center physician services in Posio is not the largest. It is among the largest, but there are areas where the coverage is higher or at the same level as in Posio. Because of this finding, the distribution of such variables as the number of visits per visitor to health center physicians and the proportion of visitors among all those who visited who made seven or more visits to health center physicians in 1975 were studied.

Table 4 shows, among other things, that when in the city of Vaasa there were an average of 1.7 visits/visitor to health center physicians; this figure in Posio was 3.6. When the proportion of the high utilizers was 1.0 per cent in Vaasa, it was 13.2 per cent in Posio. Table 4 also suggests that in all those health center areas where the average health center physician visit rate/

TABLE 4. Specific Characteristics of Health Center Physician Visits by Health Centers

Large Health Centers	Visits per Visitor	Proportion Visitors with Seven or More Visits	Small Health Centers	Visits per Visitor	Proportion Visitors with Seven or More Visits
Helsinki	2.3	4.3	Alajärvi	2.7	6.6
Oulu	1.9	2.3	Jalasjärvi	3.0	9.0
Turku	2.1	3.7	Närpiö	3.0	9.4
Tampere	1.9	2.3	Perhonjokilaakso	2.4	4.4
Kuopio	2.2	3.7	Länsi-Turunmaa	2.5	5.5
Jyväskylä	2.2	3.5	Kiihtelys-Pyhäselkä	2.3	3.5
Joensuu	2.3	3.6	Keski-Kainuu	2.7	6.5
Pori	2.1	2.8	Ilomantsi-Tuupovaara	2.7	6.8
Savonlinnanseutu	2.3	3.9	Keski-Karjala	3.3	11.1
Mikkelinseutu	2.5	5.3	Simo-Kuivaniemi	2.1	2.4
Kotka	2.1	2.7	Sodankylä	2.9	8.3
Vaasa	1.7	1.0	Posio	3.6	13.2
Rovaniemi	2.2	3.0	Ähtäarinjärvi	2.2	3.0
Lappeenranta	2.3	4.0	Kristiinanseutu	2.6	6.3
Hämeenlinnanseutu	2.3	4.0	Tuusula	2.8	7.3
Kajaani	2.7	5.9	Inari-Utsjoki	2.9	8.6
Kokkolanseutu	2.2	3.6	Nurmijärvi	2.8	6.8
Vantaa	2.3	4.3	Lamminseutu	2.6	5.5
Kemi	2.7	8.2	Kangasniemi	3.2	10.6
$\overline{X}$	2.2		$\overline{X}$	2.8	
SD	0.2		SD	0.4	
C (%)	9.1		C (%)	14.3	

visitor is 3.0 or more, the proportion of those visitors who made seven or more visits is about 10 per cent or more. The finding also seems to correlate well with the visit rates in Table 2. This finding seems to suggest that the high utilization rates result from the fact that the same persons visit frequently. The differences between the big and small health center areas are clear but the differences are also considerable within the group of the small health centers.

In order to see whether this explanation is valid we studied a group of 16 health centers, eight of which represent the lowest average health center physician use rates per person in 1975 in the country, and the other eight the highest utilization rates.

In Table 5, the health center physician population ratios are also included. From the figures in the table it can be concluded that there seems to be a relatively good correlation between the number of visits per visitor and the proportion of high utilizers among these visitors. (There also seems to be a good correlation between these two variables and the supply of health center physicians.

The findings thus indicated that:

1) there are differences in utilization patterns of ambulatory physician services between urban and rural areas. When only the public ambulatory physician service use rates (HC physician and outpatient department services which both are controlled by the national planning) are studied the use per population seems to be higher in rural than in urban areas ($p < 0.02$, t-test), whereas looking at the total use of illness-related ambulatory physician services (private sector use added), the populations of the urban areas seemed to use more services ($p < 0.001$, t-test).

2) When variations within the groups were studied it was shown that there were relatively large variations in both groups and that for some reason the variability seemed to be greater within the group of the rural health centers.

3) In light of the findings it can be stated that the national health planning mechanism which was put into effect in 1972 seems to have succeeded in leveling off the great differences in utilization of health services existing before 1972 between the urban and rural areas. This was one of the main goals of the 1972 legislation and the reorganization of the health services in Finland based on it.

TABLE 5. Health Center Physician Visits per Visitor in Health Centers Representing the Lowest and the Highest Rates in Finland 1975

Centers with Lowest Rates				Centers with Highest Rates			
Health Center	Visits per Visitor		High Utilizers (%)	Health Center	Visits per Visitor		High Utilizers (%)
Vaasa	1.7	(1.0)	1.0	Ikaalinen	3.9	(3.6)	17.4
Pietarsaari	1.8	(1.3)	1.2	Janakkala	3.7	(2.6)	15.0
Seinäjoki	1.8	(1.4)	1.2	Hämeenkyrö	3.7	(3.5)	14.4
Mustasaari	1.9	(1.2)	1.4	Posio	3.6	(4.1)	13.2
Lahti	1.9	(1.8)	1.6	Ruovesi	3.5	(4.4)	13.4
Oulu	1.9	(2.1)	2.3	Hanko	3.5	(3.9)	12.9
Tampere	1.9	(1.6)	2.3	Parkano	3.4	(3.5)	12.8
Kauniainen	1.9	(1.5)	2.7	Harjavalta	3.4	(2.9)	12.2
$\overline{X}$	1.9	(1.6)	1.7	$\overline{X}$	3.6	(3.6)	13.9

Discussion

According to the findings presented there were considerable differences in use of physician services between the different types of health center areas in Finland in 1975 by type of services used. In large population centers, considerably more services were used per person. There seemed to be an especially great difference in use of private physician services—use of specialist services—despite the universal sickness insurance. In this respect, people living in the rural areas seemed to be in an inferior position. It is not certain on the basis of the data available, however, whether people in the big population centers use specialists for true specialist medical service needs or whether these specialists are used for first-line primary care due to the lack of ordinary primary care physicians (Fig. 2). In case the demand in cities reflects a real need for specialist care (which is not probable) then one of the priorities should obviously be the increase of the availability of specialist services also to the populations of the rural areas. An effective method for doing this in Finland would be the development of the availability and accessibility of the outpatient services of general hospitals in the developing regions. This is because it is impossible to force private physicians to these regions but it is possible to allocate positions to general hospitals located in these regions, and it is also possible to designate the positions for the outpatient departments. This would improve the possibilities of consultation services for health centers. Taking into account the development in the number of physicians in the country as well as the fact that the great majority of the physicians are salaried staff of public institutions, it is almost certain that in a relatively near future even specialists will go where the positions are. It is understandable from the figures that looking only at the utilization rates for health center physicians one may get an

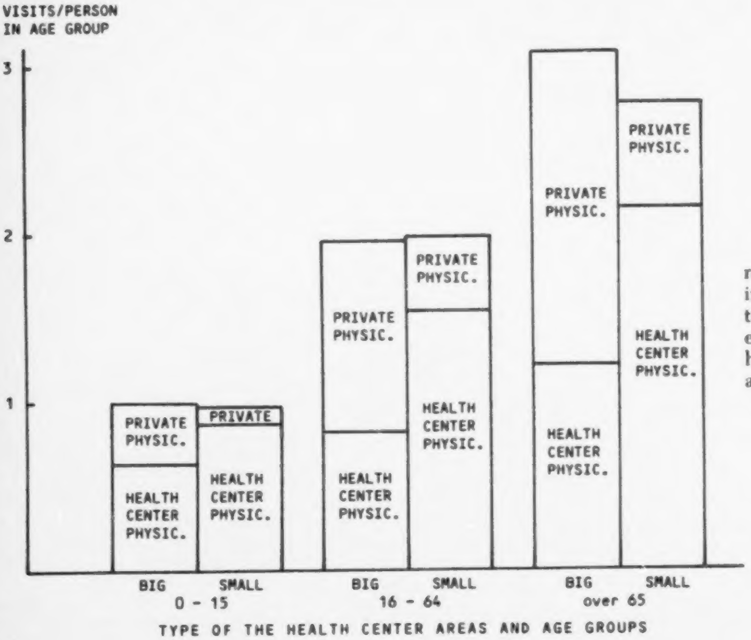


FIG. 2. Use of illness-related physician services in ambulatory care (hospital outpatient depts excluded) by type of the health center area and by age groups in 1975.

impression that the populations of the large urban health centers are in a worse position from the equity point of view. Worse here referring to the availability of cheap (to the patient) health center physician services. As mentioned earlier, the original out-of-pocket cost of a physician visit in 1975 was about 7.5 times less expensive in the health center than in the private sector. In case lab tests or x-ray examinations were ordered the difference was much greater.

Thus, the basic problem in the big centers seems not to be the lack of services but 1) the skewness of the structure of the supply which, perhaps, results in use of specialist services when the demand would be satisfied by general practitioners if enough were made available; 2) the great difference of the out-of-pocket costs of the services in the two sectors—public and private. It is likely that when the choice is between "free" health center physician services and private physician services, for which the patient has to pay considerably more, the demand is for health center services and the population considers its position worse in the big centers because the low-cost services are not as readily available as in the rural areas.

In the small health centers the structure of the supply is skewed toward another direction: there are basic services available but there may be barriers for use of specialist services. When the average visit rates per visitor were calculated, differences were also found within the groups. Especially notable, these differences were among the small health centers. The great differences probably are partly due to differences in treatment habits and partly also due to differences in available resources between the various health centers. Where the visiting rate was high, without exception there was also a large proportion of high utilizers to be found. This was confirmed by examining the 16 health centers with the highest and lowest health center physician utilization rates per visitor in the country. The observed high utilization rates per visitor and the large proportions of the high utilizers connected with them can at least partly be explained by the great number of repeat visits (whether needed or not) in these health centers. The basic reasons for these differences cannot, however, be explained by the data available, since more information would be needed on the consulting practice, the patient mix, organization of resources, etc. from these health centers. These findings, however, stress the importance of examining the effectiveness and efficiency aspects of the health services. On the basis of the figures presented, it cannot be stated what can be considered an excessively high visiting rate or what the rate should be. It would be, however, interesting for further analyses to be able to distinguish between the first and repeat visits, which is, at the moment, impossible. Also it would be interesting to study why the variability in use (and probably also in supply) of health center physician services is so great even within the groups.

Finally, a notable finding pointing to a favorable impact of national health planning can be seen in Figure 3, which shows a remarkable equality between the two groups of health center areas, as for the public services which are controlled by planning. The general impression from the findings is that by 1975 the coverage of the basic physician services in the rural areas in Finland was relatively good. However, although the health center services in big cities in general seemed to be undeveloped, there is no question about the clustering of services in them, which also shows in the bigger total utilization rates. However, based on the strong demand for the development of the health centers in the bigger urban areas existing at present, it seems that we are now approaching a point where a serious consideration should take place as to how to pro-

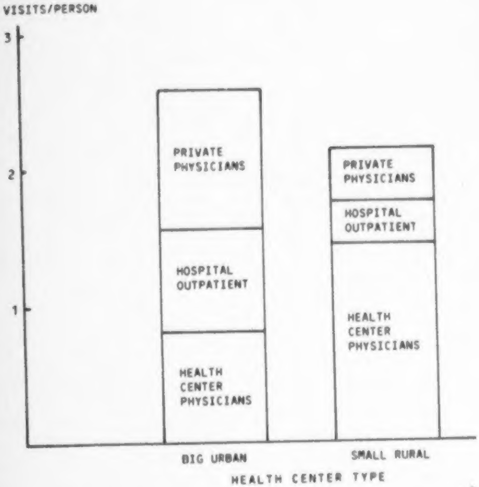


FIG. 3. Use of illness-related physician services in ambulatory care by type of the health center area in 1975.

ceed. The underlying dilemma is how to avoid increasing the inequality in resource supply and availability in favor of the urban areas again. One of the basic problems being linked with the difference in pricing of the health center and private sector services. The characteristics of the two service sectors are also different (general practitioners who are salaried versus specialists working on a fee-for-service basis who, at the same time, are salaried full-time hospital physicians). So, we are not comparing like with like. At this point it is impossible to tell what the development will be. The decisive factor will be the taste of the consumer.

Acknowledgment

The author wishes to express his thanks to Dr. Robert L. Parker of the Department of International Health at Johns Hopkins School of Hygiene and Public Health for his comments on the original manuscript.

References

1. Griffiths DAT. Inequalities and management in the N.H.S. Hospital 1971;229.

2. Hart TJ. The inverse care law. Lancet 1971;i:405.

3. Purola T, Kalimo E, Nyman K, Sievers K. National health insurance in Finland: its impact and evaluation. Int J Health Serv 1973;3:69.

4. Kekki P. Same goals-different philosophy. Br Med J 1976;1:204.

5. Kekki P. General practice in Finland. J R Coll Gen Pract 1976;26:853.

6. Tuomikoski H. Terveyspalvelusten tarjonta ja tavoitettavuus Suomessa 1964–1976. (The Supply and Availability of Health Services in Finland 1964-76.) Thesis. Dept. of Social Policy, University of Helsinki, November 1978.

7. Kekki P. Analysis of relationships between availability of resources and use of health services in Finland. A cross-sectional study. Med Care 1980;18:1228.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Visits to Physicians by Asian/Pacific Americans

ELENA S. H. YU, PH.D.,* AND BEULAH K. CYPRESS, PH.D.†

The purpose of this paper is to disseminate some of the findings from the 1979 National Ambulatory Medical Care Survey (NAMCS) where, for the first time in its history, data are available on the characteristics of visits to office-based physicians by Asian/Pacific Americans. The NAMCS Program is the only source for such data which is based on a national probability sample survey of physician–patient encounters. Among the findings are: 1) no significant difference was observed between Asian/Pacific Islanders and the white majority with regard to the patient's sex, return-visit ratio, or the kinds of diagnostic and therapeutic services rendered by the sampled physicians. 2) Children appear to be the major utilizers of ambulatory medical care among Asian/Pacific Americans. 3) In every age group, the visit rate to office-based physicians is lower for Asian/Pacific Americans when compared with white Americans. 4) A significantly smaller percentage of Asians or Pacific Islanders, in contrast to other specified race/ethnic groups, had visited a physician for injury or poisoning, and a substantial proportion of their visits were apparently made for preventive care. 5) Significantly fewer visits were made by Asian/Pacific Americans to the office of a surgeon or a psychiatrist. The demographic antecedents and cultural underpinnings of these findings are discussed.

BETWEEN 1970 AND 1980, the total number of Asians and Pacific Islanders in the United States more than doubled, from 1.5 to 3.5 million.[1] This dramatic increase came about largely through legal immigration made possible by the 1965 Amendments in the U.S. immigration law (which took effect in 1968) and the passage of the Refugee Act following the fall of South Vietnam and other parts of IndoChina. The surge in Asian/Pacific Islander population during the 1970s has exceeded that found for all other ethnic groups including persons of Spanish origin. If the present trend continues, even greater increase in the Asian/Pacific American population is expected by the 1990s. Yet, little is known about the health status, services utilization, and ambulatory medical care of this growing minority.

In 1979, however, the National Ambulatory Medical Care Survey (NAMCS) conducted by the National Center for Health Statistics (NCHS) allowed for the identification of "Asian/Pacific Islanders" on the questionnaire filled out by a probability sample of physicians selected for the survey in the coterminous United States. Thus, for the first time in the history of the NAMCS Program, data are available for minorities other than black Americans re-

---

* Division of Analysis, National Center for Health Statistics, Department of Health and Human Services, Hyattsville, Maryland.

† Division of Health Care Statistics, National Center for Health Statistics, Department of Health and Human Services, Hyattsville, Maryland.

Address for reprints: Elena S. H. Yu, Ph.D., Division of Analysis, National Center for Health Statistics, FCB#3, Room 2-27, 3700 East West Highway, Hyattsville, Maryland 20782.

0025-7079/82/0800/0809/$01.40 © J. B. Lippincott Co.

garding the characteristics of their visits to office-based physicians. Although there are methodologic problems in the interpretation of "Asian/Pacific Americans" because the label actually encompasses a heterogeneous population from diverse cultures,‡ the importance of obtaining at least some preliminary baseline information on this ethnic group seems to override the disadvantages. To date, there are no systematic studies on the ambulatory medical care of Asian and Pacific Americans. The NAMCS data appear to be the only source of information for such type of care, and it is the only data set which is based on a national probability sample survey of physician–patient encounters.

The purpose of this paper is to disseminate some of the findings from the NAMCS Program where the number of Asian/Pacific Americans for selected characteristics are sufficiently large as to permit meaningful tabulations. Data for black and white Americans are also included in the presentation that follows for the purpose of comparison. It is hoped that the general availability of these data, however limited, will stimulate greater interest among investigators in the field of health services to conduct more systematic research on the health care of minorities.

Sample Design of NAMCS

The target population of NAMCS encompasses office visits within the coterminous United States made by ambulatory patients to nonfederally employed physicians who are principally engaged in office practice, but not in the specialities of anesthesiology, pathology, or radiology. Also excluded were physicians principally engaged in teaching, research, or administration. Likewise, telephone contacts and nonoffice visits were not included in the scope of the survey. A multistage probability sampling design is used which involves probability samples of primary sampling units (PSUs), physician practices within PSUs, and patient visits within practices. For 1979, a sample of 3,023 non-Federal, office-based physicians was selected from master files maintained by the American Medical Association and American Osteopathic Association. The physician response rate was 71.8 per cent. Sampled physicians were asked to complete a questionnaire called Patient Records for a systematic random sample of office visits taking place within their practice during a randomly assigned weekly reporting period. A total of 45,351 Patient Records representing an estimated 556,313,431 office visits were completed by sampled physicians.[2] Of these, 442 records, representing an estimated 5,559,524 visits, characterized patients as "Asian/Pacific Islanders."

Limitations of the NAMCS Data

Statistics from the 1979 National Ambulatory Medical Care Survey were derived by a multistage estimation procedure, which produces essentially unbiased national estimates[2] and has three basic components: 1) inflation by reciprocals of the probabilities of selection, 2) adjustment for nonresponse, and 3) a ratio adjustment to fixed totals. Notwithstanding these safeguards, caution is warranted in the interpretation of the NAMCS data for Asian/Pacific Americans. First, ethnic identification is not made by the patients themselves but by the sampled physicians or their assistants, based on their prior knowledge of patients or observations. Second, differences exist between Asian/Pacific American subgroups and within each national origin grouping with regard to their time of first arrival in the United States, reasons for immigration, socioeconomic status, and educational attainment, to mention only a few factors.[3,4] However, insofar as Asian/Pacific Americans tend to be viewed as "alike" and are treated accordingly in the health care setting, it is legitimate for some data analysis purposes to examine their health utilization behavior as a group. Third, due to the nature of the sampling design and the resultant disparity in sample sizes between specified race/ethnic groups, the estimates of the volume and characteristics of visits to office-based physicians are more precise for the majority white population than those obtained for Asian/Pacific Americans. Despite this caution, the NAMCS data for Asians and Pacific Islanders are sufficiently precise as to be useful in two important senses: 1) by way of contrast with similar data obtained from the white majority, and 2) by providing a preliminary estimate of the characteristics of the minorities sampled in the NAMCS Program.

---

‡ Broadly speaking, the term Asian/Pacific Americans designates residents of the United States from the following countries and territories: *East Asia*: China (including Taiwan and Hong Kong); Japan and Korea. *Southeast Asia*: Philippines, Vietnam, Cambodia, Laos, Thailand, Malaysia, Singapore, and Indonesia. *Indian Subcontinent or South Asia*: India, Pakistan, Bangladesh, Sri Lanka, and Burma. *Pacific Islands*: Hawaii, Guam, Samoa, Tonga, Fiji, and other Micronesian Islands. In practice, however, only a few major groups originating from the Asian-Pacific triangle are identified in federal statistics. These major groups include persons identified as Chinese, Japanese, Filipinos, Koreans, IndoChinese (Vietnamese, Cambodians, and Laotians), Hawaiians, Guamanians, and Samoans. Some Census Bureau reports have also identified (East) Indians separately.

NAMCS Data for Asian/Pacific Americans

In 1979, the majority of the physician visits made by Asian/Pacific Americans occurred in metropolitan areas (93.8 per cent) rather than in nonmetropolitan areas (6.2 per cent). In contrast, less than three-quarters of white American patient visits occurred in the metropolitan areas (Table 1). This pattern of concentration found for Asians and Pacific Islanders is consistent with their geographic distribution in the United States. For instance, an independent estimate based on the National Health Interview Survey (NHIS) conducted in 1978 by the National Center for Health Statistics indicates that nearly 92 per cent of Asian/Pacific Americans reside in standard metropolitan statistical areas.

In the NAMCS data, no discernible difference was observed between Asian/Pacific Americans and the white majority with regard to the patient's sex. In both groups, proportionately more females (60 per cent) than males (40 per cent) visited physicians during the survey period. In general, 45 per cent of the visits by Asian/Pacific Islanders and 38 per cent of those by white Americans were for new problems. The ratio of return visits to new

TABLE 1. Per Cent Office-based Physician Visits in Metropolitan and Nonmetropolitan Areas for Specified Race/Ethnic Group: United States, 1979

	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent	100.0	100.0	100.0	100.0
Metropolitan area†	73.4	72.5	81.8	93.8
Nonmetropolitan area	26.6	27.5	18.3	6.2

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the Coterminous United States).

† Standard metropolitan statistical area as defined by the U.S. Bureau of the Census.

TABLE 2. Illness-related Visiting Rates/1000 in Various Ambulatory Care Sectors by Health Center Area

Large Health Centers	HC-Physician Visits	Private Physician Visits	Outpatient Visits	Total	Small Health Centers	HC-Physician Visits	Private Physician Visits	Outpatient Visits	Total
Helsinki	739	1458	1046	3243	Alajärvi	1321	668	297	2286
Oulu	493	1153	1069	2715	Jalasjärvi	1773	179	237	2189
Turku	361	1648	1291	3300	Närpiö	1545	210	284	1914
Tampere	477	1235	767	2479	Perhonjokilaakso	1177	320	382	1879
Kuopio	738	1305	674	2717	Länsi-Turunmaa	1142	816	312	2270
Jyväskylä	756	1026	533	2315	Kiihtelys-Pyhäselkä	1185	245	665	2095
Joensuu	1021	603	801	2425	Keski-Kainuu	1391	291	362	2044
Pori	564	1555	628	2747	Ilomantsi-Tuupovaara	1506	118	384	2008
Savonlinnanseutu	1062	589	531	2182	Keski-Karjala	1988	248	118	2354
Mikkelinseutu	1107	638	570	2360	Simo-Kuivaniemi	715	381	579	1675
Kotka	812	1070	700	2582	Sodankylä	1746	122	110	1978
Vaasa	378	1063	839	2280	Posio	2122	160	132	2414
Rovaniemi	910	776	612	2298	Ähtärinjärvi	887	674	342	1903
Lappeenranta	996	779	746	2518	Kristiinanseutu	733	773	471	1977
Hämeenlinnanseutu	862	1084	572	2518	Tuusula	1490	629	471	2590
Kajaani	1206	688	714	2608	Inari-Utsjoki	1703	122	126	1951
Kokkolanseutu	742	876	890	2508	Nurmijärvi	1550	654	415	2619
Vantaa	925	1257	613	2795	Lamminseutu	1439	342	236	2017
Kemi	1145	564	783	2492	Kangasniemi	1780	271	220	2271
$\overline{X}$	805	1019	757	2583	$\overline{X}$	1431	380	323	2128
SD	259	337	201	295	SD	154	57	24	251
C (%)	32.2	33.1	26.6	11.4	C (%)	10.8	15.0	7.4	11.8

TABLE 3. Health Center and Private Physician Visit Rates per 1000 by Age Groups in 1975

Large Health Centers	HC-Physician Visit			Private Physician Visit			Small Health Centers	HC-Physician Visit			Private Physician Visit		
	0–15	16–64	65–	0–15	16–64	65–		0–15	16–64	65–	0–15	16–64	65–
Helsinki	607	676	1275	631	1570	2005	Alajärvi	765	1445	1911	124	773	1349
Oulu	360	497	948	472	1284	2368	Jalasjärvi	948	1972	2246	40	235	154
Turku	440	256	842	829	1722	2629	Närpiö	861	1556	2345	81	236	264
Tampere	487	452	595	430	1331	2171	Perhonjokilaakso	572	1243	2336	61	412	399
Kuopio	581	765	945	414	1412	2822	Länsi-Turunmaa	1028	1085	1515	302	862	1284
Jyväskylä	645	744	1151	415	1125	1839	Kiihtelys-Pyhäselkä	573	1259	2091	58	296	359
Joensuu	624	1079	1485	150	673	1286	Keski-Kainuu	709	1542	2165	63	371	350
Pori	503	560	732	512	1814	2281	Ilomantsi-Tuupovaara	872	1601	2184	34	144	139
Savonlinnanseutu	723	1112	1508	129	657	1198	Keski-Karjala	1072	2105	2985	101	281	343
Mikkelinseutu	738	1132	1733	177	755	1318	Simo-Kuivaniemi	355	817	1143	118	486	502
Kotka	584	833	1116	440	1079	2205	Sodankylä	1058	1964	2585	23	164	148
Vassa	430	348	460	271	1182	2049	Posio	1127	2425	3641	33	217	207
Rovaniemi	688	967	1208	190	921	1631	Ähtärinjärvi	743	897	1096	122	755	1284
Lappeenranta	738	1034	1268	253	868	1319	Kristiinanseutu	413	733	1171	208	808	1410
Hämeenlinnanseutu	674	865	1184	367	1179	1859	Tuusula	1043	1560	2385	263	779	568
Kajaani	836	1273	1883	116	843	1260	Inari-Utsjoki	1227	1835	2319	28	163	124
Kokkolanseutu	504	771	1313	272	1028	1693	Nurmijärvi	1076	1663	2166	291	790	743
Vantaa	855	895	1787	646	1465	1852	Lamminseutu	932	1454	2022	109	405	387
Kemi	782	1208	1637	266	612	1000	Kangasniemi	974	1896	2399	47	344	277
$\overline{X}$	621	814	1241	367	1133	1831	$\overline{X}$	860	1529	2142	111	449	542
SD	141	294	400	195	359	513	SD	246	452	619	90	258	448

TABLE 2. Per Cent Office-based Physician Visits by Type of Therapeutic Services Rendered for Specified Race/Ethnic Group: United States, 1979

Therapeutic Services Rendered	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent†	100.0	100.0	100.0	100.0
None	19.8	20.2	15.5	18.0
drugs (prescription and non-prescription)	51.3	50.1	62.5	54.5
Injection	9.6	9.4	11.3	13.1
Immunization/ desensitization	5.2	5.4	3.2	6.3
Diet counseling	6.0	5.8	7.7	6.9
Family planning	1.4	1.4	2.3	1.8
Medical counseling	22.2	22.3	22.1	19.1
Physiotherapy	3.1	3.0	3.7	2.9
Office surgery	7.4	7.6	4.8	7.6
Psychotherapy/therapeutic listening	4.4	4.6	3.1	2.5
Other therapy	3.5	3.6	2.1	2.8

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the United States).

† Per cent do not add to total because more than one service may be rendered per physician visit.

problem visits was 1.2 for Asian/Pacific Americans and 1.6 for white Americans, a difference which is not statistically significant.† Likewise, in terms of the kinds of diagnostic and therapeutic services rendered by the sampled physicians, so significant differences across specified race/ethnic groups were observed (Tables 2 and 3). Thus, Table 2 shows that the proportion of Asians and Pacific Islanders who received immunizations or desensitizations (6 per cent) was not significantly different from that observed for white Americans (5.4 per cent). Despite the stereotyped belief that Asians have a strong sense of modesty, an equal proportion of Asian and Pacific Islander women (5 per cent) received Pap tests as did white and black Americans (Table 3). Insofar as blood pressure check-ups are concerned, significantly more black Americans (44.5 per cent) had actually received this diagnostic service than did Asian/Pacific Islanders (33 per cent) or white Americans (35 per cent).

There are, however, significant differences between the specified racial/ethnic groups insofar as the age distributions of the patients are concerned (Table 4). For Asian/Pacific Americans, as much as 30.4 per cent (roughly 1.7 million divided by

---

† The significance of the difference between two percentages was tested using the t-test with a critical value of 1.96 (0.05 level of significance) according to the following formula:

$$t = \frac{p_x - p_w}{\sqrt{s^2 p_x + s^2 p_w}}$$

Since NAMCS is based on a complex sampling design, standard errors cannot be developed according to formulas based on the assumption of random sampling. The half-sample replication procedure was used to calculate standard errors of estimates in the survey.[5-6] Approximate standard errors of percentages based on 1979 aggregate estimates may be calculated by the following formula:

$$SE_p = \sqrt{\frac{30.96199 \times p \times (1 - p)}{X}}$$

TABLE 3. Per Cent Office-based Physician Visits by Type of Diagnostic Services Rendered for Specified Race/Ethnic Group: United States, 1979

Diagnostic Services Rendered	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent	100.0	100.0	100.0	100.0
None	10.2	10.6	6.0	8.2
Limited exam/history	63.0	63.0	63.1	64.8
General exam/history	16.8	16.6	18.5	20.3
Pap test	4.9	5.0	4.7	5.0
Clinical lab test	23.2	23.0	25.3	21.8
X-ray	8.2	8.2	8.6	5.2
Blood pressure check	36.0	35.3	44.5	33.2
EKG	2.7	2.7	3.1	0.5
Vision test	6.0	6.2	4.5	6.6
Endoscopy	1.3	1.4	0.8	1.8
Mental status exam	1.5	1.5	1.3	0.4
Other diagnostic service	3.5	3.6	2.5	3.9

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the coterminous United States).

† Per cent do not add up to total because more than one service may be rendered per physician visit.

5,559,524) of estimated visits to physician's office were made by children under 15 years of age, which is significantly different from that observed for white or black Americans (18.2 per cent and 17.1 per cent, respectively). Using the population estimates obtained from the National Health Interview Survey by N.C.H.S. in 1979, we have calculated the visit rate for the different racial/ethnic groups. The findings are shown also in Table 4. First, compared with white or black Americans, Asian/Pacific Americans have the lowest visit rate to office-based physicians. Second, while

TABLE 4. Estimates and Rates of Visits to Office-based Physicians by Age for Specified Race/Ethnic Group: United States, 1979

	All Races		White		Black		Asian/Pacific	
Age Group	Visit Estimates	Visit Rate*	Visit Estimates	Visit Rate*	Visit Estimates	Visit Rate*	Visit Estimates	Visit Rate*
All ages	556,313,431	2.58	502,926,839	2.70	46,789,200	1.84	5,559,524	1.55
Under 15 years	101,352,298	2.02	91,267,680	2.21	8,015,492	1.03	1,690,800	1.98
15–24 years	82,289,782	2.05	74,295,285	2.18	7,209,758	1.38	675,243	0.99
25–44 years	151,713,912	2.59	133,337,485	2.61	16,485,828	2.58	1,590,530	1.49
45–64 years	128,594,299	2.96	117,859,164	3.04	9,550,199	2.35	1,061,339	1.51
65 years and over	92,363,140	3.96	86,167,225	4.08	5,527,923	2.71	541,612	1.94

* Based on estimates of the civilian, noninstitutionalized population of the United States excluding Hawaii and Alaska, obtained from the Health Interview Survey, 1979. In the HIS Program, race for black and white Americans was based on interviewer's report. Race of Asian/Pacific Americans was based on respondent's self-report.

the visit rate for white and black Americans shows an almost linear increase with age, that for Asian/Pacific Americans displays a more U-shaped relationship with age. Third, within-group comparison indicates that among Asian/Pacific Americans, in contrast to the majority of Americans, the youngest age group have the highest visit rate.

In terms of medical specialty, about one-third of the visits for both white and Asian/Pacific Americans were made to the offices of General and Family Practitioners. However, significantly more visits by Asians and Pacific Islanders (19 per cent) compared with white majority (10 per cent) or black minority (10.6 per cent), were made to the offices of pediatricians. These data are, of course, consistent with the earlier finding that a sizable proportion of the physician visits by Asian/Pacific Islanders were made by patients under 15 years of age. Dermatology is another specialty where a statistically significant difference exists between Asians and white Americans. In addition, a significantly smaller percentage of the visits by Asians and Pacific Islanders vis-a-vis white majority were made to the office of a general or specialty surgeon or a psychiatrist (Table 5). The cultural underpinnings of these findings will be discussed in the following section of this paper.

Table 6 presents data on physician visits by major ICD-9 Diagnostic Codes[7] for specified race/ethnic groups. Only 1.9 per cent of the visits made by Asian/Pacific Americans compared with 4.5 per cent for the white majority were for mental disorders. The difference is statistically significant. In contrast, significantly more visits made by Asian/Pacific Islanders (10.2 per cent) than by white Americans (5.3 per cent) were for problems related to diseases of the skin and subcutaneous tissue. Examination of the more detailed ICD-9 codes reveals a clue to the underlying reasons for this peculiar "skin problem" among Asian/Pacific Islanders. Contact dermatitis and other eczemas, as well as diseases of sebaceous glands are two diagnoses in which the per cent visits are proportionally higher among Asian/Pacific Americans, compared with white and

TABLE 5. Per Cent Office Visits by Physician Specialty for Specified Race/Ethnic Group: United States, 1979

Medical Specialty	Total* U.S.	White	Black	Asian/ Pacific Americans
Total				
estimate of visits	556,313,431	502,926,839	46,789,200	5,559,524
per cent	100.0	100.0	100.0	100.0
General and family practice	34.2	33.4	42.3	32.0
Internal medicine	12.0	12.3	9.3	9.3
Pediatrics	10.5	10.3	10.6	19.3
Obstetrics and gynecology	9.1	9.0	10.2	13.5
Dermatology	3.2	3.3	1.2	7.1
Other medical specialities	3.8	4.0	2.7	1.9
General surgery	6.1	6.1	6.5	3.8
Other surgical specialty	16.0	16.3	13.4	11.7
Psychiatry	3.1	3.2	1.7	0.6
Other	2.1	2.1	2.5	0.8

* Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the coterminous United States).

black Americans. Since these diagnoses include diaper rash and acne, their relative preponderance is most likely an artifact of the overrepresentation of children under 15 years old among Asians and Pacific Islanders.

Table 6 also shows that 23 per cent of the Asian or Pacific American visits compared with only 16 per cent for the white majority were made for the category "factors influencing health status and contact with health services" (which used to be called "special conditions and examinations without sickness"). This diagnostic group includes general medical examinations, routine normal pregnancy (prenatal) examinations, and supervision of healthy children (well-baby examinations), which together constitute the largest proportion of visits for specific diagnoses regardless of race. Of further interest is the significantly smaller proportion of injury and poisoning (which was called "accidents, poisoning, and violence" in the ICD-8) observed for Asian and Pacific Islanders (5.3 per cent) compared with white (9.3 per cent) or black (9.6 per cent) Americans. The implications of these findings are explored below.

Cultural Diversity Reflected in NAMCS Data

In many ways, the results of our analyses are both heartening and disheartening. On one hand, the visit rate for Asian/Pacific Americans obtained from the NAMCS data is rather low in every age group except for those under 15 years, compared with that found for white and black Americans. A perusal of the National Health Interview Survey data at NCHS lends an interesting interpretation to the NAMCS finding. From 1976 through 1978, visits to office-based physicians on the average accounted for 66.5 per cent of the total annual encounters by Asian/Pacific Islanders with sources of medical care, compared with 69.3 per cent for white Americans (NCHS, unpublished HIS data).‡ In contrast, visits to emergency rooms and/or outpatient clinics represented as much as 18.8 per cent of the total visits by Asian/Pacific Americans, compared with 11.7 per cent for white Americans. In terms of the rate of visits to emergency rooms and/or outpatient clinics per 1,000 population, Asian/Pacific Americans had a higher rate—798.1§ (annual average from 1976–78 pooled data)—than white Americans in 1978 (566.0§) but a lower rate than black Americans (1,188.8§ for 1978). These NHIS findings suggest that the NAMCS data give only a partial picture of the use of health services by Asian/Pacific Americans. It is most likely that language and cultural barriers have prevented many immigrant Asians and Pacific Islanders from having a regular "family" doctor, as had been reported earlier in a few localized studies.[8,9] Lack of health insurance, ignorance about the service delivery system in the United States, and a sizable number of unattached adults and elderly in the Asian populations are three other possible factors in the heavy reliance of immigrants on emergency room and/or outpatient clinics for medical care.

On the other hand, at least one reason for optimism exists. Taken at face value, the NAMCS data fail to show gross disparities—across specified race/ethnic groups—in the kinds of diagnostic and therapeutic services rendered by the sampled physicians to their patients. Instead, some interesting cultural diversity in medical care is reflected in the NAMCS data. For instance, within the Asian/Pacific Islander population, children under 15 years of age have the highest rate of visit to

---

‡ Prior to 1977, racial breakdowns in the HIS tabulations did not routinely allow for the identification of black Americans separately.

§ The visit rate is age-adjusted by the direct method to the 1970 civilian noninstitutionalized population, using four age intervals: Under 17 years, 17–44 years, 45–64 years, and 65 and over.

TABLE 6. Per Cent Office-based Physician Visits for Specified Race by Principal Diagnoses and ICDA Code: United States, 1979

Principal Diagnosis and ICD-9-CM Codes*		Total* U.S.	White	Black	Asian/ Pacific Americans
All classes					
estimate of visits		556,313,431	502,926,839	46,789,200	5,559,524
per cent		100.0	100.0	100.0	100.0
Infectious and Parasitic Diseases	(001–139)	3.5	3.5	3.4	4.1
Neoplasms	(140–239)	2.6	2.7	1.6	1.4
Endocrine, nutritional, and metabolic diseases, and immunity disorders	(240–279)	4.1	4.0	5.1	3.2
Mental disorders	(290–319)	4.4	4.5	4.1	1.9
Diseases of the nervous system and sense organs	(320–389)	9.1	9.4	6.3	9.1
Diseases of the circulatory system	(390–459)	8.9	8.8	10.6	7.3
Diseases of the respiratory system	(460–519)	13.2	13.1	13.6	15.1
Diseases of the digestive system	(520–579)	4.4	4.5	3.9	6.6
Diseases of the genitourinary system	(580–629)	6.6	6.3	9.9	5.0
Diseases of skin and subcutaneous tissue	(680–709)	5.2	5.3	3.8	10.2
Diseases of the musculoskeletal system and connective tissue	(710–739)	6.7	6.5	8.3	4.0
Symptoms, signs, and ill-defined conditions	(780–799)	3.1	3.1	2.9	1.8
Injury and poisoning	(800–999)	9.3	9.3	9.6	5.3
Factors influencing health status and contact with health services	(V01–V82)	15.8	15.9	14.2	23.3
Other diagnoses		1.5	1.5	1.6	0.6
None or unknown		1.6	1.6	1.1	1.1

* Based on the Ninth Revision, International Classification of Diseases, Clinical Modification, September 1980.

† Includes data for other ethnic groups not shown separately (i.e., Indians and Alaskans in the coterminous United States).

‡ Includes diseases of the blood and blood-forming organs, complications of pregnancy, childbirth and the puerperium, congenital anomalies, certain causes of perinatal morbidity and mortality, blank diagnosis; noncodable diagnosis; and illegible diagnosis.

office-based physicians, compared with the other age groups. We suspect that this may be due to the presence of a sizable proportion of persons in the under 15 age group because Asian/Pacific Americans are, by and large, still very much an immigrant population. This is borne out by an examination of the U.S. Census data on the proportion of foreign-born persons for different race/ethnic groups. For instance, while only five per cent of the white population in 1970 was born outside of the

United States, the lowest percentage of foreign-born persons within the five *major* Asian American subgroups[‖]—21 per cent—is found among Japanese Americans.[10] An inspection of thę annual reports prepared by the Immigration and Naturalization Service[11] reveals that the influx from the Philippines, Korea, China (including Hong Kong and Taiwan), and IndoChina (Vietnam, Laos, Cambodia)—in contrast to Japan—has increased dramatically as a result of the relaxation of U.S. immigration restrictions in 1965. Hence, the foreign-born segment of the Chinese population (47 per cent as of 1970 Census) and that of the Filipino population (53 per cent as of 1970) is likely to increase dramatically when the 1980 Census tabulations are released within the next two years. Although corresponding figures for Koreans and IndoChinese are not available from the 1970 Census, the immigration records provide evidence that the majority of these two Asian subgroups will be foreign-born by virtue of their sudden surge in the 1970s. Some researchers have estimated that it took the IndoChinese less than five years to reach the population size in the U.S. comparable to that achieved by the Chinese in over 100 years of piecemeal immigration—a conjecture which remains to be tested against the 1980 Census results. Thus, we can expect that the majority of the Asian Americans will continue to retain many of their traditional values and health practices from the old countries. Accordingly, health is viewed as a state of equilibrium ("homeostasis") between man, society, and the cosmic forces of the universe. Maintaining a balance between the "hot" and "cold" elements of the body is the cornerstone of good health. Disease is a disturbance of that relationship. Blood is a source of human vitality and difficult to replenish. Consequently, surgery is avoided to the extent possible since it increases the risk of losing one's blood. These concepts, of course, are so deeply rooted in the traditional cultures of developing societies that they crisscross ethnic boundries. For various Asian ethnic groups, however, some of the similarities in herbal prescriptions (e.g., the use of Ginseng for restoring vitality or stress endurance) and behavioral proscriptions (e.g., never sleeping with wet hair; "doing the month" after delivering a baby) are rather striking.[12–16] Taking into consideration the major composition of the Asian/Pacific American population, i.e., Chinese, Japanese, Korean, IndoChinese, and Filipino, it is apparent that historically at least four of these groups have been heavily influenced by China—in language, religion, philosophy, social structure, and medicine. Such pervasive influence was made possible by the diffusion and use of Chinese characters as the primary vehicle of written communication and the medium of instruction among the literati in ancient Japan (5th to 19th century), Korea (108 B.C.–1886), and Vietnam (111 B.C.–16th century). Although the Filipinos had their unique alphabet system, the fact that Chinese presence in the Philippines antedates the Spanish Conquest (began in the 16th century) by at least 500 years, also suggests the possibility of Chinese influence in Filipino society—though not to the same extent as that found in other countries neighboring China. Therefore, despite the apparent cultural diversity among Asians, unifying threads of social values and health beliefs are interwoven into the indigenous cultures. The persistence of these health-related beliefs and practices in Asian American subcultures is evident from a few studies in the United States based on limited samples and sometimes impressionistic observations.[17–19] Obviously, precise empirical support for the prevalence and sociodemographic variations of these ethnic practices are lacking.

[‖] Chinese, Japanese, Filipino, Koreans, and Vietnamese.

In the NAMCS data, demographic factors probably played an important role in the high physician-visit rate of Asian/Pacific Americans for "special conditions and examinations *without* sickness," while cultural factors may be salient in their unusually low proportion of injury and poisoning (or what was called "accidents, poisoning, and violence" in the ICD-8). The former is essentially *preventive* health care, while the latter, though usually unanticipated, is *preventable* in most cases. Thus, the juxtaposition of these two findings would tend to suggest that although proportionally fewer Asian/Pacific Americans rely upon office-based physicians for their regular source of medical care, many of those who do visit physicians are apparently prevention-oriented. It is further possible that this consciousness for the importance of prevention may extend beyond the supervision of infants.

As early as 1966, Kurokawa[20,21] noted that "Oriental" children in California had had far fewer accidents than Caucasian children. She hypothesized that this may be due to the peculiar family structure found among Asians (primarily Japanese and Chinese) where adults other than the child's parents are not only present, but also participate in the supervision and discipline of the young. In contrast, Caucasian families are mostly nuclear in structure, and parental rights are seldom fully shared with other adult members, even when they are present. To further prove her point, Kurokawa interviewed the mothers of 151 Oriental children who were enrolled in the Kaiser prepaid medical plan in California. She found that acculturated (i.e., "Americanized") Oriental children had more accidents than the unacculturated, because the former were more likely to be independent and venturesome which, in turn, exposed them to potential hazards, whereas the latter tended to be "oversupervised" by American standard. Findings from other independent studies indicate that traditional Asian cultures indeed reinforce dependency behavior in the child, which restricts his ability to explore the unfamiliar all by himself.[22-24] American child socialization, on the other hand, places great emphasis on independence training.[25,26] Hence, it is possible that, relatively speaking, American child socialization exposes the young to greater risks of injury due to accidents, poisoning, and violence. Nevertheless, the small sample size in Kurokawa's study makes broad generalizations from her findings difficult. Regardless of this limitation, her work does provide some food for thought and demonstrate the dire need for more systematic studies on the family structure and health practices of Asian/Pacific Americans.

Another interesting cultural pattern manifested in the National Ambulatory Medical Care Survey data is the significantly small percentage of visits made by Asian/Pacific Americans to the office of a psychiatrist. Moreover, regardless of the medical specialty of the physician they visited, only a very small percentage of Asian/Pacific American patients received the principal diagnosis of "mental disorders." Thus, the NAMCS data lend confidence to previous observations made by practicing psychiatrists and clinical psychologists. To illustrate, in a study of 17 community mental health services in the greater Seattle area, Sue[27] found that over a period of three years, only 3.1 per cent of the Asian American clients saw a psychiatrist at intake, and not one Asian saw a psychiatrist during therapy. A number of researchers believe that cultural factors play an important role in the asian reluctance to consult with psychiatrists, even when they admit to having psychological problems. For the most part, Asians and Pacific Islanders appear to have a tendency to somaticize their psychiatric symptoms.[28] Hence, the consensus among practitioners is that members of this minority group generally perceive "talk" therapy to be ineffective.[29,30] Instead, self-medication, such as herbal tonics, are the expected treatment even for mental problems, thereby causing an extensive delay in help-seeking until the disorder has clearly become unbearable or unmanageable. Lin and Lin[31] reported from their experience that the delay among Chinese patients in seeking psychiatric help can take as long as 25 years. Intrafamilial resources are first utilized before outside help is sought and, even then, respected intermediaries who have the family's trust are consulted before assistance is sought from a psychiatric professional who is a total stranger to the family. These findings, consistent as they are, are not based on large-scale epidemiologic studies. Hence, until prevalence and incidence data on both "treated" and "untreated" mental disorders among Asian/Pacific Americans are available, the extremely low percentage of visits to psychiatrists by this growing minority must be interpreted with caution.

Conclusion

Taking into consideration the limitations of the data collected from the National Ambulatory Medical Care Survey conducted by the National Center for Health Statistics, we have presented the major findings on physician visits by Asian/Pacific Americans. On one hand, no significant difference was observed between Asian/Pacific Islanders and the white majority with regard to the patient's sex, return-visit ratio, or the kinds of diagnostic and therapeutic services rendered by the sampled physicians. On the other hand, some interesting cultural differences are discernible from the patterns of visits to office-based physicians reported for Asian/Pacific Islanders. Children, followed by the elderly, appear to be the major utilizers of ambulatory medical care among Asian/Pacific Americans. However, in almost every age group, the visit rate to office-based physicians is lowest for Asian/Pacific Americans, compared with either black or white Americans. When compared with white Americans, a significantly small percentage of Asians or Pacific Islanders had visited a physician for injury and poisoning; and a substantial proportion of their visits were apparently made for preventive care—general medical examinations, routine normal pregnancy examinations, and supervision of healthy children. Significantly few visits were made by Asian and Pacific Americans to the office of a surgeon or a psychiatrist. A cultural resistance to the utilization of these two types of medical specialty is suggested.

Acknowledgments

The authors thank their colleagues Edward Bacon (Director of the Division of Health Care Statistics, N.C.H.S.), Joel Kleinman (Director of the Division of Analysis, N.C.H.S.), William T. Liu (Director of the Pacific/Asian American Mental Health Research Center and Professor of Sociology, University of Illinois at Chicago Circle), and two anonymous reviewers for *Medical Care*, for their constructive criticisms of an earlier draft of this paper. Another colleague, Thomas Drury (Acting Chief, Health Status Measurement Branch), was most helpful in sharing some of the unpublished HIS data referred to in this paper.

References

1. Bureau of the Census. 1980 Census Population Totals for Racial and Spanish Origin Groups in U.S. Announced by Census Bureau. Washington, D.C.: United States Department of Commerce, Public Information Office, Feb. 23, 1981.

2. National Center for Health Statistics. 1979 summary ambulatory medical care survey. Advancedata 1981;66:1.

3. Yu E. A sociodemographic profile of Asian Americans, 1820–1976. Paper presented at the Ethnic Resources Institute of Rosary College, River Forest, Illinois, October 19–20, 1979.

4. Yu E, Drury T, Liu WT. Estimating the health status of Asian/Pacific Americans: conceptual and methodological problems. Papers prepared for the Proceedings of the Annual Meeting of the American Statistical Association, Detroit, Michigan, August 1981.

5. McCarthy PJ. Replication: an approach to the

analysis of data from complex surveys. Vital and Health Statistics. Series 2, No. 14. DHEW Pub. No. (HSM) 73-1269. National Center for Health Statistics, Health Services and Mental Health Administration. Washington: U.S. Government Printing Office, April 1966.

6. McCarthy PJ. Pseudoreplication: further evaluation and application of the balanced half-sample technique. Vital and Health Statistics, Series 2, No. 31. DHEW No. (HSM) 73-1270. National Center for Health Statistics, Health Services and Mental Health Administration. Washington: U.S. Government Printing Office, January 1969.

7. World Health Organization. International Classification of Diseases. Manual of the international statistical classification of diseases, injuries, and causes of death, vol I. Geneva: W.H.O., 1980.

8. Li F, Schlief NY, Chang CJ, Gaw AC. Health care fo the Chinese community in Boston. Am J Public Health 1972; 62:636.

9. Chinatown, U.S.A., 1970. California Health 1970;2:1.

10. Sung BL. Chinese american manpower and employment. Washington, D.C.: U.S. Department of Labor Manpower Administration, 1975.

11. Immigration and Naturalization Service. Immigration and Naturalization Reports, 1943–1979. Washington, D.C.: U.S. Government Printing Office.

12. McKenzie JL, Chrisman NJ. Healing herbs, gods, and magic: folk health beliefs among Filipino-Americans. Nurs Outlook 1977;25:326.

13. Gould-Martin K. Hot cold clean poison and dirt: Chinese folk medical categories. Soc Sci Med 1978;12:39.

14. Popov IM, Goldwag WJ. A review of the properties and clinical effects of ginseng. Am J Chinese Med 1973;1:263.

15. Sich D. Traditional concepts and customs on pregnancy, birth and post partum period in rural Korea. Soc Sci Med 1981;15B:65.

16. Pillsbury, BLK. "Doing the Month": confinement and convalescence of Chinese woman after childbirth. Soc Sci Med 1978;12:11.

17. Ling S, King J, Leung V. Diet, growth, and cultural food habits in Chinese-American infants. Am J Chinese Med 1975;3:125.

18. Campbell T, Chang B. Health care of the Chinese in America. Nurs Outlook 1973;21:245.

19. Chan CW, Chang JK. The role of Chinese medicine in New York City's Chinatown. Am J Chinese Med 1976;4:129.

20. Kurokawa M. Family solidarity, social change, and childhood accidents. J Marriage Fam 1966;28:498.

21. Kurokawa M. Acculturation and childhood accidents among Chinese and Japanese Americans. Genet Psychol Monogr 1969;79:89.

22. Yu E, Liu WT. Fertility and Kinship in the Philippines. Notre Dame: University of Notre Dame Press, 1980.

23. Young NF. Independence training from a cross-cultural perspective. Am Anthropologist 1972;74(3):629.

24. Vogel E, Vogel S. Family security, personal immaturity, and emotional health in a Japanese sample. J Marriage Fam 1961;23:161.

25. Winterbottom M. The Relation of Childhood Training in Independence to Achievement Motivation. M. A. thesis, University of Michigan, 1953.

26. Kriger SF, Kroes WH. Child rearing attitudes of Chinese, Jewish, and Protestant mothers. J Soc Psychol 1972;86:205.

27. Sue S. Community mental health services to minority groups. Some optimism, some pessimism. Am Psychologist 1977;32:616.

28. Tseng WS, McDermott JF Jr, Maretzki TW, eds. People and Cultures in Hawaii: An Introduction for Mental Health Workers. Honolulu: Department of Psychiatry, University of Hawaii School of Medicine, 1974.

29. Chien C-P, Yamamoto J. Effective psychotherapy for the Asian American and Pacific Islander patients. In: Acosta F, et al., eds. Psychotherapy for Low Income and Minority Patients. Prepublication copy, 1981.

30. Tung TM. The IndoChineses Refugees as Patients. Refugee Resettlement Exchange Project. Washington, D.C.: American Public Welfare Association, 1980.

31. Lin T-Y, Lin M-C. Service delivery issues in Asian–North American communities. Am J Psychiatr 1978;135:454.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Drug Product Selection in Illinois: The First Year's Experience

JESSE E. STEWART, PH.D., PAUL G. GRUSSING, PH.D.,
AND ANAL A. PUROHIT, PH.D.

This paper reports an evaluation of Illinois' drug product selection (DPS) law. The results show that while consumers saved an average of $1.47 whenever DPS occurred in the study period, the prescription format mandated in the Illinois law appears to be a hindrance to unambiguous communication from prescribers. In addition, because of the positive formulary approach adopted by Illinois, the potential for DPS behavior was limited to only 20.8 per cent of the prescriptions audited in the study and ultimately, DPS occurred in 12.6 per cent of this sample. It also is reported that DPS was more likely to occur in urban community pharmacies and that consumers saved more money when DPS occurred in this environment, but prices for prescriptions dispensed as written were also higher in this environment. Finally, it was found that prescription prices were affected by the type of pharmacy patronized, but were not affected by the urban or rural location of pharmacy.

ALTHOUGH DRUGS PRESENTLY account for less than 10 per cent of the total cost of medical care in the United States, they have been identified as representing an area of potentially substantial savings to consumers, for a variety of reasons. Proponents of prescribing and dispensing of drugs by generic (rather than brand) name, such as the Federal Trade Commission and the Department of Health and Human Services through the Food and Drug Administration, argue that a considerable portion of consumer expenditures on prescription drugs could be saved if the practice of selecting less expensive generic drug products was encouraged.[1,2]

For the 20 years before 1970, virtually all states had laws which specifically forbade the pharmacist to dispense a drug manufactured or distributed by any company other than the brand-name drug called for on the prescription. Only if the prescription was written generically could the pharmacist dispense any manufacturer's equivalent product in the pharmacy. In the early 1970s, the states began repealing these so-called antisubstitution laws, however, replacing them with a variety of laws aimed at saving consumers' money.

These drug product selection (DPS) laws allow the pharmacist, usually in specified instances, to dispense a less expensive generically equivalent drug product in lieu of the branded product originally prescribed. A variety of mandatory

---

From the Department of Pharmacy Practice, University of Illinois, College of Pharmacy, Chicago, Illinois.

Supported in part by the Illinois Department of Public Health, Springfield, Illinois. Data coding and computerization was performed by the Survey Research Laboratory, University of Illinois, Chicago Circle Campus. The paper represents part of a larger research effort which is being reported in several independent communications. Special thanks are expressed to Mr. Ronald Gottrich who as a Graduate Research Associate was invaluable in the successful completion of this project.

Address for reprints: Jesse E. Stewart, Ph.D., University of Illinois, College of Pharmacy, 833 South Wood Street, Room 142, Chicago, IL 60612.

0025-7079/82/0800/0831/$01.40 © J. B. Lippincott Co.

and permissive clauses in the various laws have produced an environment in which the laws can basically be classified in three categories. The first and largest category includes those laws which do not require a formulary,* but simply allow the pharmacist to dispense a less expensive, generically equivalent product, unless prohibited by the prescriber or the patient. The second category, representing almost as many states, includes those laws which mandate the development of a positive formulary. Thus, when pharmacists select drug products, they can only do so within the limitations of a formulary, which generally lists those multiple-source drugs judged to be equivalent (either therapeutically or chemically, depending upon state law) by a committee of experts. The last category includes those states enacting laws that require the development of a negative formulary. In these states, the pharmacist may use drug product selection for any drug unless it is listed in this formulary or prohibited by the prescriber. The DPS laws in 28 states now require the development of a positive or negative formulary. A variety of additional clauses, such as prescription signature format, prevention of DPS by the prescriber, cost savings pass-on requirements and patient consent requirements add to the variability of laws from state to state and have been accurately summarized elsewhere.[3]

Drug product selection laws have now been enacted in almost every state, but evidence of how effective the various laws are is still limited. Only a few studies have evaluated the effectiveness of the DPS laws through the use of actual prescription data. While some[4] might argue that shopping surveys represent the most appropriate methodologic approach, the expense involved in using this approach generally has forced researchers to review only a small number of drugs in a limited number of pharmacies. Thus such results are limited in their applicability to other geographic areas or even to other drugs or drug classes.

The first major evaluation of DPS legislation was conducted by Goldberg and colleagues in Michigan, using Wisconsin as a control state.[5–7] The Michigan law, which is permissive by not mandating DPS, does not use a formulary and allows prescribers to prohibit DPS by stating so on the prescription. By examining prescription data for the year prior to enactment of Michigan's law and for the subsequent 2 years, they found that significant savings occurred when DPS took place ($1.23 per prescription in year 2) as well as when prescriptions were written generically. The savings were less dramatic, however, when the prescription initially was written generically. They also reported that the rate of DPS did not increase substantially from the first year of the law to the second, and was found to be only 1.50 per cent of all prescriptions for multiple source products.

A second major study recently was reported by Vuturo and colleagues following their evaluation of Florida's DPS law.[8] A unique feature of the Florida law requires the state to develop a negative formulary and each pharmacy to develop and display its own positive formulary for use when selecting drug products. The researchers reported that an average savings of $1.92 was achieved each time DPS occurred in their study sample, and despite the formulary limitations of the Florida law, the DPS rate of 6.2 per cent for multiple-source prescriptions compared favorably with the 1.5 per cent reported in Michigan and the 2.7 per cent reported in Wisconsin.

The present study represents the third major study examining the effectiveness of a state's DPS law that used actual prescription data collected on a statewide basis and was not restricted to a limited number of drugs for comparison. On July 1, 1978, the Illinois Drug Product Selection law became effective and incorporated a number of features found in the laws of other states. The law amended the Illinois Food, Drug and Cosmetic Act to allow DPS if the product is listed in a positive formulary to be developed, maintained and issued by the Illinois Department of Public Health. The remaining features of Illinois' law were enacted as amendments to the Illinois Pharmacy Practice Act and are highlighted below:

---

* A list of drug entities showing pharmaceutically or therapeutically equivalent products within each entity. A formulary includes both branded products and those marketed only under a generic name.

1. The prescription blank must contain two signature lines, one indicating "may substitute," the other indicating "may not substitute," and two accompanying check boxes.
2. DPS may occur when the "may substitute" option is properly signed and checked or when the prescriber fails to properly check and sign beside one of the alternatives.
3. When presented with a prescription where drug product selection can be used, the pharmacist may inform the patient of the availability of a less expensive product and the price of the alternate product.
4. The person for whom the prescription is written, a spouse, or a guardian can provide oral permission to the pharmacist allowing DPS.
5. The alternate product must be less expensive than the product that was originally called for in the prescription.
6. When DPS occurs, the pharmacist must record the name and address or other identification of the manufacturer of the product ultimately dispensed.
7. The act of DPS by a pharmacist will not constitute evidence of negligence if the alternate drug is included in the Illinois Drug Formulary.

These seven areas represent the legislative intent as well as the mechanics of the Illinois law. In addition, a unique feature of the law states that "the Department of Public Health shall have the duty of monitoring the cost savings effected by substitution . . . and shall annually report to the legislature as to the cost savings being achieved." A study was, thus, designed to answer this legislative mandate as well as to examine several other important features of the law in order to provide evidence of the law's effectiveness. That portion of the study described here had three primary objectives:

1. Measure the rate of DPS in Illinois according to pharmacy type, payment and geographical variables.
2. Identify certain factors which preclude or promote DPS behavior by physicians and pharmacists.
3. Determine cost savings to consumers as a result of the DPS law.

Methodology

To answer the evaluation questions posited in this research, the prescription was identified as the basic element or primary unit of investigation. The study population was thus defined as all new prescriptions dispensed between May 1, 1978, and June 30, 1979, in Illinois pharmacies that were subject to the Drug Product Selection law.†

A two-stage sampling design was utilized. The first stage sampling involved selection of pharmacies and identified geographic location and type of pharmacy as important selection variables. All licensed pharmacies in the state were divided into urban and rural categories as defined by Standard Metropolitan Statistical Areas. Next, all pharmacies were classified as either community, chain (four or more locations), or hospitals providing outpatient pharmacy services using acceptable definitions. Clinic pharmacies were combined with community pharmacies and a total of 227 other pharmacies, such as nursing home pharmacies and nuclear pharmacies, not affected by the law, were excluded from the study.

---

† The 14-month study period represented the first 12 months after the law had been enacted (July 1, 1978–1979) plus the two-month period prior to enactment for data baseline purposes. All data reported are based on the 12-month study period only.

Initially, 60 pharmacies were selected systematically with a random start from the list of licensed pharmacies generated previously. The sample size for each of the six strata (community, chain and hospital pharmacies versus urban or rural location) was selected proportional to the number of pharmacies in each stratum. The rural chain cell and hospital outpatient cell were oversampled choosing six pharmacies in each cell since initial selection of one pharmacy in each cell was deemed to be too small for comparison purposes.‡ In addition, because of the disproportionately small number of hospitals providing outpatient pharmacy services, the urban and rural hospital cells were collapsed. Thus, data were ultimately collected from 70 systematically selected pharmacies representing the three types of pharmacies in both urban and rural environments, with the hospital cell representing both environments.

The second stage of the sample selection involved identification of new prescriptions dispensed during the 14-month study period. Due to the expected low rate of DPS, a large number of prescriptions needed to be examined in order to answer the evaluation questions. However, because most prescriptions were expected to be dispensed as written (DAW), it was not necessary to generate a proportionately large sample of prescriptions to determine the prescribing and dispensing characteristics of physicians and pharmacists. To accommodate these extremes and to facilitate comparison with previous studies, the methodology of auditing and reviewing prescriptions used in previous studies was adopted.[7]

On the average, 200 clusters of 10 prescriptions were surveyed from the 70 pharmacies. The first prescription of the cluster of 10 was an "audited" prescription and the remaining nine prescriptions of the cluster were considered "reviewed" prescriptions. The purpose of the audited prescription was to collect detailed data about the prescription. The purpose of the reviewed prescription was to enlarge the sample of drug-product-selected prescriptions without necessarily enlarging the sample of DAW prescriptions, since the DPS rate, the unit of study, was expected to be low. For the reviewed prescriptions, such detailed data were collected only if drug product selection was used for the prescription. Using this methodology, a theoretical total of 140,000 prescriptions would be examined visually (200 × 1 × 70 audited prescriptions + 200 × 9 × 70 reviewed prescriptions).

To select prescriptions proportional to pharmacy prescription volume, each of the 70 selected pharmacies was telephoned and asked to provide their prescription volume for the 14-month study period.§ Prescription volumes then were determined separately for each of the five cells, and selection intervals for the clusters of 10 prescriptions were determined for each cell independently. The sample of prescription clusters were then selected proportional to the prescription volume in each pharmacy. Thus, larger volume pharmacies had more prescriptions audited and reviewed than did smaller pharmacies (i.e., more clusters selected).

Data were collected by the investigators as well as by graduate students in pharmacy administration. A one-day training session was completed by all data collectors which included actual use of the data collection form. On the average, the error rate of data collectors was less than three per cent, and subsequent to actual data collection, each individual was retrained in those areas where errors had occurred. Thus, the error rate in the field was expected to be even lower.

---

‡ Appropriate weighting procedures were subsequently employed so that the statistical analysis of data would reflect the number of pharmacies accurately representing each cell.

§ Each of the reported volumes were later verified by data collectors in each pharmacy and no volume or sampling adjustments were subsequently required.

Results and Discussion

The primary intent of DPS legislation is to decrease the amount of money that consumers must spend on prescription drugs. The following discussion will examine: 1) the potential market for DPS in Illinois, 2) the rate of DPS in Illinois, and 3) the savings occurring when the pharmacist practiced DPS.

Potential for Drug Product Selection

Drug Product Selection can only occur in Illinois if the prescription is written using a product's brand name, if that drug entity[‖] is listed as interchangeable in the *Illinois Drug Product Selection Formulary* and if both the prescriber and patient allow the use of an alternate product. While each of these factors was incorporated into the law to protect the consumer and the professional prerogative of the prescriber, each tends to limit the potential instances of DPS.

Earlier studies in Michigan[6] and Florida[8] noted that just over 50 per cent of the prescriptions examined were multiple source and, thus, potentially eligible for DPS. The remainder of the prescriptions were single source products, most of which were still under patent protection. In Illinois, even when a prescription drug is multiple source, it must be written using a

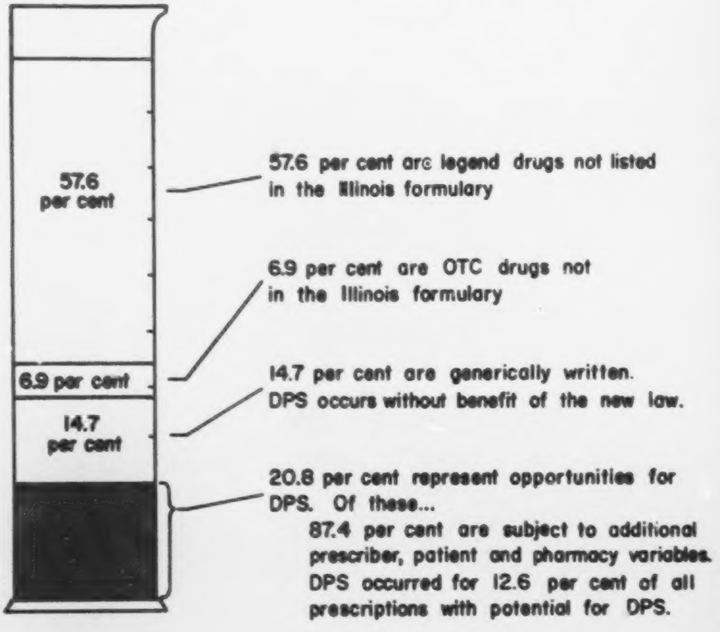


FIG. 1. Opportunities for drug product selection. Data represent audited prescriptions (N = 12,071) for the first 12 months of Illinois' DPS law.

[‖] Drug entity—a unique drug with individual characteristics and specifications. Several equivalent brands or specific drug products may occur for the same drug entity.

brand name and be listed in the Illinois Formulary before DPS can be even considered. Those prescriptions written using the drug's generic name may be dispensed by the pharmacist with either a brand name product or a less expensive generically marketed product, just as occurred prior to this law.

Figure 1 graphically displays the opportunities for DPS within the confines of the Illinois law. Of the 12,071 prescriptions audited in the 12-month study period, 57.6 per cent were legend¶ drugs not listed as interchangeable in the Illinois Formulary and are referred to as nonformulary. Another 14.7 per cent (12.3 legend and 2.4 per cent over-the-counter) were written generically and were not affected by this law. Products not requiring a prescription (over-the-counter) are excluded from the Illinois law also. These represented 6.9 per cent of all audited prescriptions. Finally, it can be seen that only 20.8 per cent of all prescriptions audited in the study were written using the drug's brand name and were listed in the Illinois Formulary. This represents the potential area of opportunity for the pharmacist to practice DPS, given physician and patient consent. The rate or occurrence of DPS is most appropriately expressed based on this category alone.

Rate of DPS

An examination of those prescriptions that were legend, brand written, and listed in the formulary (20.8 per cent of all audited prescriptions) indicates that DPS occurred 12.6 per cent of the time. These results compare favorably to data reported for Florida (14.2 per cent)[8] and Vermont (14.4 per cent),[9] but are in marked contrast to the very low rate reported for Michigan (1.5 per cent).[7] A complete discussion of comparison of DPS experiences across those few states reporting such data is beyond the scope of this paper, but the reader is cautioned that differences in state laws, research methodologies and means of reporting the data prevent direct comparisons unless a thorough understanding of these differences is achieved first.

In order to examine whether certain identifiable environmental factors affect the rate of DPS in Illinois, an analysis of rates was conducted for each of the five major pharmacy types studied. Table 1 indicates that urban community pharmacies showed the highest rate of DPS (15.3 per cent) while hospital pharmacies showed the lowest rate for their outpatient dispensing activities. The high rate for urban community pharmacies is explained at least partly by higher proportional numbers of Illinois Public Aid (Medicaid) prescriptions dispensed in this environment with that program effecting de facto DPS for financial reasons. The low rate (five per cent) reported in the hospital outpatient environment reflects a high rate of generic prescribing in the environment (18.3 per cent) combined with the existence of a hospital formulary which in itself restricts prescribers to a limited number of brand name products which they learn to prescribe.

In addition to differences in DPS rates between pharmacy types, urban versus rural environments also produced notable differences. The rate of DPS for all urban pharmacies studied was 14.3 per cent while DPS was much less likely to occur in rural pharmacies (7.0 per cent rate).

While the DPS rate in Illinois is dependent upon the fact that the prescription is written using a brand name, and that the drug entity is listed in the Formulary it also is dependent upon the physician's signature and approval format. As mentioned, the Illinois law mandates prescribers to

---

¶ According to federal law, a "legend" drug requires a prescription from a licensed practitioner and the following legend must appear on the manufacturer's label: "Caution: Federal law prohibits dispensing without prescription."

TABLE 1. Rates of DPS for Audited Prescriptions that were Legend, Brand Written and Formulary*

	Type of Pharmacy									
	Urban Community		Rural Community		Urban Chain		Rural Chain		Hospital Outpatient	
Pharmacist Behavior	N	%	N	%	N	%	N	%	N	%
Prescriptions drug product selected	145	15.3	27	6.33	43	11.78	21	9.10	10	5.00
Prescriptions dispensed as written	795	83.95	396	92.96	320	87.67	210	90.90	190	95.00

* Columns may not equal 100% because of rounding and because of instances of DPS outside the Formulary.

use a prescription blank which contains two signature lines and accompanying check boxes, one indicating "may substitute," the other indicating "may not substitute." Approximately 26 states have provision for the use of a two-line prescription format, but few, if any, besides Illinois, require the accompanying check boxes.** Because of this unique feature of the law, the effect of clear, unclear, confusing, and contradictory prescription signature formats was also examined.

Although the Illinois DPS law went into effect on July 1, 1978, the state allowed prescribers an additional, unspecified amount of time beyond that date to obtain prescription blanks containing the required 2-signature format. In the present study, when DPS occurred, only 36.6 per cent of the prescriptions examined were in format compliance on the effective date of the law (July 1, 1978). The rate of compliance rose, however, to 76.8 per cent by the end of the first year. Alternately, for those prescriptions that were dispensed as written, only 33.9 per cent were in compliance on the effective date of the law and this rose to 64.4 per cent after one year. Of all audited prescriptions, 55.25 per cent were found to be in compliance with the legal prescription format required. These data suggest that DPS occurred more frequently when the prescription format was in legal compliance.

Except in those cases where the prescriber has clearly indicated through signing and checking the accompanying box that DPS is allowed or prohibited, the intent of the prescriber may not be accurately communicated to the pharmacist under the Illinois law. If the prescriber signs either line but fails to check the boxes, signs one line but checks the alternate box, or continues to sign the bottom of the prescription as was done in the past, the pharmacist may be confused as to the prescriber's intent. Thus, the researchers created the labels of clear, unclear, confusing and contradictory to describe and further analyze the various signature formats encountered.

For all instances in which DPS occurred, the prescriber's instructions were in clear compliance only 9.9 per cent of the time. The remaining prescriptions were either unclear, confusing or contradictory in their instructions to the pharmacist, were telephoned prescriptions, or drug products were selected based upon the state's Public Aid program or on the basis of prior consent. Prescribers who either favor or oppose DPS could achieve their desired

---

** As of October 1980, 21 states and the District of Columbia required the physician to write a supplemental phrase on the prescription form such as "dispense as written—DAW," "No sub," "NS," "brand necessary" or "medically necessary" in order to prohibit drug product selection.

objectives by communicating more clearly to the pharmacist. The problem of lacking clarity is not a simple matter of prescriber intent but is more clearly a problem of the complex prescription format requirements that are on impediment to both the prescriber and the pharmacist.

Cost Savings

Of primary concern to consumers, legislators and health practitioners is the question of whether consumers saved money when drug product selection was used for prescriptions during the first year of the DPS law. Table 2 shows that significant savings occurred when DPS was practiced in Illinois. Further, it indicates that for private-pay prescriptions, an average of $1.30 was saved when drug product selection was used for a prescription, and that the state saved $1.88 on Public Aid prescriptions in the same way. The difference between the mean price for all three categories of the dispensed-as-written prescriptions and those where drug product selection was used (Table 2) shows that an average of $1.47 was saved every time DPS occurred. This average savings is consistent with savings reported from other states which range from $0.86 to $1.92 per prescription.[6,8] Again, notable differences in various state DPS laws prohibit an in-depth interpretation of these differences in savings in this paper.

A second methodology was employed to validate the reported savings. When the mean price for all products for which drug product selection was used is compared with the mean price for all products which were dispensed as written, an additional variable may contribute to the savings reported. If the drugs which were dispensed as written were largely more expensive drugs on the market, and if the prescrip-

TABLE 2. Analysis of Variance (ANOVA): Average Prescription Price by Payment Mechanism*

	Dispensing Behavior				
	Drug Product Selection (DPS)		Dispensed as Written (DAW)		
Payment Mechanism	Mean	N	Mean	N	Savings
Private pay	5.00	(1118)	6.30	(1750)	1.30
Public aid	4.60	(1220)	6.48	(281)	1.88
Other third party	8.14	(53)	6.72	(110)	−1.42
Total	4.87	(2401)	6.34	(2156)	1.47

Source	Degrees of Freedom	F Value
Dispensing Behavior (DB)	1	8.37†
DB at private pay	1	105.38†
DB at public aid	1	73.79†
DB at other third party	1	6.59†
Payment Mechanism (PM)	2	36.51†
PM at DPS	2	30.82†
PM at DAW	2	1.10
Dispensing Behavior by Payment Mechanism	2	25.17†
Within Cells	4526	

* Included audited and reviewed prescriptions which are legend, brand written and Formulary.

† $p \leq 0.01$

tions for which DPS occurred were generally for less expensive drugs, then the effect of this mix would confound the findings. To validate the reported savings of $1.47 per prescription, a modified matched-pair technique was used to ensure that the comparisons were based on the same drug entities.†† This technique showed that the savings for all entities for which matched-pairs were located was $1.45. This, compared to the finding of $1.47, served to validate the findings based on a comparison of mean prices for dispensed-as-written prescriptions versus those for which drug product selection was used.

The larger savings occurring with Public Aid prescriptions reflects two direct pricing influences upon the pharmacist. When a drug entity is single source, the Public Aid program reimburses the pharmacist an amount based on the Estimated Acquisition Cost (EAC) for that product. However, for those drugs that are multiple source, the pharmacist is reimbursed based on a Maximum Allowable Cost (MAC) established at the federal or state level which essentially reimburses for a lower priced generically equivalent drug, unless approval is obtained from the Department of Public Aid. In addition, the pharmacist may only dispense a generically equivalent product when so authorized by the prescriber, even though reimbursement for the drug product is fixed at the lower price. Such authorization occurs most often in the form of blanket authorization‡‡ from the prescriber for Public Aid patients. Thus, Public Aid policy in conjunction with a lowest-cost reimbursement approach virtually mandates that pharmacists use drug product selection for these prescriptions and assures the state of greater savings in drug costs. Complete reimbursement from the state for Public Aid prescriptions also includes a fixed dispensing fee per prescription dispensed.

Table 2 also shows analysis of variance (ANOVA) statistics for the reported mean prescription prices under two pharmacists' dispensing behaviors and three payment mechanisms. Main effects for dispensing behavior and payment mechanism, and the interaction of these two were found to be significant ($p \leq .01$). All three dispensing behaviors indicated by simple main effects that consumers saved significantly when DPS occurred.§§ Further, the Scheffe's method was employed to identify individual mean comparisons that were significant when DPS occurred. It was found that patients whose prescriptions were paid for by a third party were charged significantly ($p \leq .01$) more when DPS occurred as compared to private-pay and Public Aid patients. A systematic review of the third party prescriptions for which

---

[††] For the last three months of the study period, when a prescription was legend, brand written and listed in the formulary, and when it was dispensed as written, the pharmacist was asked to price it as though drug product selection had been used. The pharmacist also was asked if he/she stocked the entity by another manufacturer, and if so, would he/she use drug product selection for this entity. The same procedure was carried out for those prescriptions for which drug product selection had been used and the pharmacist was then asked to price it as though it had been dispensed as written. To determine the reliability of this price estimate procedure, the pharmacist was first asked to furnish prices for 10 prescriptions that already were audited by the surveyor. These 10 estimated prices furnished by the pharmacist were compared via a paired t-test statistical method with the actual prices when the drug was dispensed. If the prices were determined to be reliable (t-test was not significant at the .05 level) the pricing procedure described above was completed. Only data from pharmacies with demonstrated reliability (consistency) in pricing were used for this matched-pairs study.

[‡‡] Blanket authorization is a delegatory act by a prescriber in which prior consent is given to a pharmacist to practice DPS.

[§§] When applied to the mean price for third party prescriptions that were dispensed as written versus drug product selected, the t-statistic for unequal variance was not significant. It was however, significant at $p \leq 0.01$ for both private pay and Public Aid prescription prices.

TABLE 3. Analysis of Variance (ANOVA): Drug Product Selection (DPS) Savings Occurring Based on Pharmacy Types*

	Urban Community		Rural Chain		Urban Chain		Hospital Outaptient		Rural Community	
Pharmacist Behavior	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
DPS Prescriptions	4.98	(1128)	4.08	(132)	4.35	(319)	5.08	(141)	5.75	(128)
Prescriptions Dispensed as Written (DAW)	6.80	(781)	5.84	(201)	5.60	(319)	6.00	(166)	6.32	(392)
Total Savings	1.82		1.77		1.22		.92		.62	

Source	Degrees of Freedom	F Value
Pharmacy Type (Type)	4	9.51‡
Type at DPS	4	6.84‡
Type at DAW	4	5.06‡
Dispensing Behavior (DB)	1	75.62‡
DB at urban community	1	133.01‡
DB at rural chain	1	20.88‡
DB at urban chain	1	20.94‡
DB at hospital	1	5.42‡
DB at rural community	1	2.63‡
Pharmacy Type by Dispensing Behavior	4	2.73†
Within Cells	3798	

* Includes audited and reviewed prescriptions which are legend, brand written and Formulary.

† $p \leq 0.05$

‡ $p \leq 0.01$

product selection was used could not fully explain this phenomenon. Various copayment and deductible features have been shown to influence both prescribing and dispensing behaviors when a third party carrier is involved, but data to confirm such an influence were not collected.

In examining additional variables that contribute to consumer drug costs, pharmacy type and pharmacy location were each identified for further analysis. It was anticipated that these variables might help to explain price variations to consumers and public policy makers. Using a 2 × 5 ANOVA procedure comparing prices by types of pharmacies, significance of the overall savings was established ($p \leq .01$) and significance between types of pharmacies was shown ($p \leq .01$). Simple main effects were used to show the significance of savings within the five types of pharmacies. When DPS occurred, and the savings were calculated for all five cells, urban community pharmacies showed the highest savings ($1.82 per prescription), ($p \leq .01$), and rural community pharmacies showed the lowest savings ($0.62). Table 3 shows the calculated savings for each of the five types of pharmacies studied.

To further contrast urban versus rural and community versus chain pharmacy prescription prices, additional Scheffe's tests were calculated.[|||] The results showed that when DPS did not occur, prescriptions

---

[|||] The mean prices used for these comparisons are not shown in Table 3. They represent data generated by combining community and chain price data in both urban and rural environments, and by combining community price data and chain price data from both environments.

dispensed in urban pharmacies were not significantly more expensive than prescriptions dispensed in rural pharmacies ($6.45 vs $6.20). Also, when DPS did not occur, prescriptions dispensed in community pharmacies were significantly more expensive on the average, than those dispensed in chain pharmacies ($6.66 vs $5.70), ($p \leq 0.05$). Similarly, when DPS did occur, the differences in prices between urban and rural pharmacies were not significantly different ($4.85 vs $4.91). Finally, when DPS did occur, prescriptions dispensed in community pharmacies were significantly more expensive than those dispensed in chain pharmacies ($5.06 vs $4.27) ($p \leq .05$). A meaningful comparison of hospital outpatient prescription prices could not be made because of the lower purchase prices hospital pharmacies may enjoy for bona fide outpatient dispensing purposes under the Nonprofit Corporations amendment to the Robinson-Patman Act and a 1976 Supreme Court interpretation.

In summary, whether DPS occurred or not, there was no significant difference in prescription prices between urban and rural pharmacies when chain and community prescription prices were combined. Alternately, whether DPS occurred or not, there was a significant difference in the average prescription prices between those dispensed in community versus chain pharmacies when urban and rural prescription prices were combined. Thus, consumers can expect to pay a variable amount for their prescriptions depending upon the type of pharmacy patronized, but the location of their pharmacy in either an urban or rural area, was not shown to affect the price of their prescriptions significantly.

Conclusions

An evaluation of Illinois' DPS law demonstrated that it enabled consumers to save, but that various factors still impede acceptance of this new role by pharmacists and physicians. It is evident that the restrictive nature of the positive formulary (listing 172 drug entities) and the prescription format requirement of the Illinois law combine to decrease the potential for DPS by pharmacists.¶¶ This legislated change in pharmacist and physician behavior is affected by a multitude of additional attitudinal and perceptual factors, some of which were analyzed as part of this evaluation study, but because of length will be reported in additional publications.

The data presented from this research show that the rate of DPS varies between types and locations of pharmacies and is affected by the mix of private pay, Public Aid and third party prescriptions dispensed. In addition, both prescriber compliance with the mandated prescription format and the clarity of instructions to the pharmacist appeared to affect the DPS behavior of the pharmacist.

Consumer cost savings on prescription drugs are occurring in Illinois and these savings appear to vary, depending upon the source of payment for the prescription. Consumers patronizing urban community pharmacies in Illinois can expect to pay more for their prescriptions, but also can expect to save more when DPS occurs. Finally, it was concluded that prescription prices were affected by the type of pharmacy patronized but were not affected by the urban or rural location of the pharmacy.

There is a continued need to conduct further research in this area to determine those features of the various DPS laws which afford consumers access to less expensive prescription drugs, while maintaining drug quality and other professional prerogatives of both prescribers and pharmacists. The difficulty of meeting the needs of consumers, health professionals and the drug manufacturers is recognized, but not always addressed in the legislative process. Hopefully, continued research in the area will facilitate problem solving in the public interest.

¶¶ An amendment to the Illinois DPS law which simplifies the prescription format by requiring one signature line accompanied by two check boxes designated "may substitute" and "may not substitute" has been enacted recently. The present study provided much of the data utilized in promoting this amendment.

References

1. Federal Trade Commission. Drug product selection. Staff report, Bureau of Consumer Protection. Washington, D.C.: Federal Trade Commission, 1979.

2. Reich RB. Drugs and medical devices policy session briefing book. Washington, D.C.: Federal Trade Commission, 1978.

3. Anonymous. Drug substitution: where does your state stand. Nat Assoc Retail Druggist J 1980;102:44.

4. Francetic I, Lasagna L, Weintraub M, Karch F. Prescription prices under the New York generic substitution law. Ann Intern Med 1980;92:419.

5. Goldberg T, Moore WE, Koontz T, et al. Evaluation of impact of drug substitution legislation. J Am Pharm Assoc 1976;NS16:66.

6. Goldberg T, Aldridge GW, DeVito CA, Vidis J, Moore W, Dickson WM. Impact of drug substitution legislation: a report of the first year's experience. J Am Pharm Assoc 1977;NS17:216.

7. Goldberg T, DeVito CA, Smith D, et al. Evaluation of economic effects of drug product selection legislation. Med Care 1979;17:411.

8. Vuturo GJ, Krischer JP, McCormick WC. Drug product selection: the Florida experience. Am J Public Health 1980;70:479.

9. Goldberg T. Comparison of key findings relevant to the effectiveness of generic drug substitution laws in four states. Presented to the Economic, Social and Administrative Sciences Section of the Academy of Pharmaceutical Sciences, American Pharmaceutical Association Annual Meeting, Washington, D.C., May, 1980.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Podiatric Surgery: The Need for a Second Opinion

MADELON LUBIN FINKEL, PH.D.,* EUGENE G. MCCARTHY, M.D., M.P.H.,† AND DAN MILLER‡

A second-opinion podiatry program, designed to contain costs and enhance the quality of care, was established in the greater New York City area by the Building Service Employees' Health Fund. All nonemergency foot surgery recommended by a podiatrist that cost over $100 was included in this program. Of the 363 individuals who had a second opinion, 82 per cent were confirmed for surgery; however, the confirmation rate among those seeing a podiatrist serving as consultant was 94.3 per cent, while the confirmation rate of those seeing an orthopedist serving as consultant was 49.5 per cent. For every condition, the podiatrists overwhelmingly confirmed the need for surgery. Conversely, the orthopedists were more conservative in their recommendations regarding surgery. The program's cost was $311,813. If a second opinion consultation wasn't required for removal of corns or ingrown toenails, the total would be much less. It seems sensible (economically and medically) to encourage or require a consultation with an orthopedist for conditions such as bunions, hammertoes and heel spurs, and not to require a second opinion consultation for superficial podiatric conditions such as corns and ingrown toenails, which are treated predominantly by the podiatrist.

THE FOOT is probably the most used and the most neglected part of the human body. Each day an average person's feet hit the ground 20,000 times with a force that is three times one's body weight. While most people begin life with two normal feet, problems can and often do, develop from abuse or neglect. Many of these problems can be prevented with proper foot care. Yet, many regard the foot as being an undramatic part of the body and regard its problems as trivial. Many prefer not to seek proper medical or podiatric attention; however, such neglect may lead to disabling conditions. Not surprisingly, the current popularity of jogging and marathon running has created a surge of complaints concerning foot, ankle and knee problems.

At least 10 million individuals suffer from soft tissue complaints (corns, calluses, bunions and nail disorders) and/or static foot deformities (clubfoot, flatfoot). Of all the things that can go wrong with the 26 bones, 107 ligaments and 19 muscles that make up the foot, none is more common than the development of corns. Corns, simply a small callus with a compact central core, result from pressure of the hard mass against underlying tissue. The best way to get rid of corns is to eliminate the original pressure or friction that caused

---

* Assistant Professor, Department of Public Health, Cornell University Medical College, New York, New York.

† Clinical Professor, Department of Public Health, Cornell University Medical College, New York, New York.

‡ Public Health Intern (formerly) and Master's Candidate, Columbia University School of Public Health, New York, New York.

Preliminary Findings were presented at the Annual Meeting of the American Federation of Clinical Research, San Francisco, California, April 1981.

Address for reprints: Madelon Lubin Finkel, Ph.D., Department of Public Health, Cornell University Medical College, 1300 York Avenue, New York, NY 10021.

0025-7079/82/0800/0862/$01.25 © J. B. Lippincott Co.

them; i.e., get better fitting shoes. Bunions, resulting from inflammation of the joint at the base of the big toe, are one of the most common afflictions of the normal adult foot. They are four times more frequent in women than in men, primarily because women's shoes increase the valgus strain on the big toe, thus irritating the condition. The underlying cause of most bunions is improper alignment of the bones of the foot and irritation by the shoe serves to aggravate the condition. Many spend thousands of dollars having a bunion corrected, only to have it recur because the faulty bone alignment remains. Ingrown toenails, too, are a major source of foot problems. Improper cutting of the nail or improper footwear can cause this condition. It is estimated that flatfoot and soft tissue complaints are not adequately cared for, either because the individual patient does not seek care or because inadequate attention is paid to these conditions once care is sought. Conversely, most static foot deformities receive professional attention.[1]

A survey of foot problems in the continental U.S. found that 74 million Americans (40 per cent of the population) reported they have foot problems.[2] Those most prone to foot trouble are women, the elderly, those with incomes below $5000, nonwhites and those residing in the South. Roughly 23 million have had surgery performed on one or both feet and 13.5 million have had some form of medical treatment for their conditions. Simple arithmetic shows that almost 38 million people wih foot problems have had neither surgery nor medical treatment.

Most surgery on the foot is elective (nonemergency) and is performed to correct deformity, relieve pain, improve function or prevent future disability. The increase in the rate of specific podiatric surgical procedures clearly reflects the difficulty people are having with their feet. From 1971 to 1978, there was a 558.8 per cent increase in the rate of foot surgery in which the bones of the foot were cut. There was a 103.9 per cent increase in the rate of joint surgery performed on the foot or toe during this same period. In comparison, the rate of all orthopedic surgery increased 36.3 per cent from 1971 to 1978 and the rate of all operations increased 24.3 per cent during this same time period.[3]

Foot problems are handled by either an orthopedic surgeon or a podiatrist. While the former is a surgeon who specializes in treating the musculoskeletal system of the body, including the foot, the podiatrist is concerned with the diagnosis and treatment of the foot and ankle. Orthopedists see a greater percentage of foot problems related to injury, as well as static foot deformity such as clubfoot, while podiatrists treat the overwhelming majority of soft tissue complaints and flatfoot. Disorders of the toenail, for example, are probably diagnosed and treated more often by podiatrists than by any other medical specialist. In most states, podiatrists are allowed by law to deal only with the foot; however, in six states they are permitted to treat up to the knee, and in another six states they may work on the foot and tendons of the leg.[4] Most podiatrists still confine themselves to work on the foot, however. The situation is analogous to that of a nonsurgical physician who technically is licensed to set a broken bone, but he is unlikely to do so.

Podiatrists have considerable patient acceptance in the United States. Roughly 7500 podiatrists engaged in patient care of some 35 million visits in 1978, a 40 per cent increase since 1970. Although the demand is increasing, the supply of podiatrists is relatively small. While there is one medical doctor for every 670 people and one dentist for every 2,000 individuals, there is only one podiatrist for every 25,000 persons.[5]

Given the extent of foot problems among the population, it is not surprising that the cost of foot care in the United States is staggering. It has been estimated that the provision of foot care (excluding home treatment using over-the-counter remedies) costs over $1 billion a year with office visits and hospital stays accounting for $450 million each. The remainder is accounted for by surgeons' fees, emergency room care and outpatient clinics. Foot-related office visits account for about two per cent of all payments to direct providers of health care, and foot-related hospital stays account for roughly one per cent of all hospital expenditures.[1] Ironically, cost containment programs have ignored podiatric services. One innovative program, designed to contain costs related to foot care, was established in the greater New York City area by the Building Service Employees' Health Fund, Local 32B-J.

Study Population

The Service Employees International Union Local 32B-J represents workers of both sexes who are employed in the maintenance, operation and upkeep of all private and public buildings (residential and commercial) in boroughs of the City of New York, except the Bronx. Those employed as superintendents working in loft and office buildings, exterminators and window cleaners are excluded specifically. The local is chartered by the Service Employees International Union under the American Federation of Labor-Congress of Industrial Organizations—600,000 members and dependents. It is composed of 120,000 members and dependents and serves as the exclusive bargaining representative for its members, assisting them to obtain adequate compensation for their labor and promoting their welfare.

Members work as elevator operators, doormen, porters and office cleaners in over 6,000 scattered worksites. Sixty per cent work the night shifts, particularly the women who work as cleaning ladies, usually without supervision. Since the tasks of maintaining and servicing buildings require that the individual to be on their feet much of the time, it is not surprising that there is a high incidence of foot problems among the membership.

Mandatory Second Opinion Podiatry Program

From 1979 through 1980, the Building Service Local 32B-J Health Fund paid over $288,000 for podiatric surgery. Almost $251,036 was paid for in-hospital surgery and $37,727 was paid for in-office surgery. The majority of in-hospital claims were for bunion removal while the majority of the in-office surgery was for removal of ingrown toenails. It was decided to require a second opinion consultation for podiatric conditions in order to possibly contain the costs of such surgery and to hopefully enhance the quality of care.

The second-opinion program for podiatric problems was established in the fall of 1979. Each member and his or her dependents is required to schedule a second opinion consultation for all nonemergency foot surgery that is recommended by a podiatrist costing over $100. Whether the surgery is to be performed on an inpatient or outpatient basis is immaterial. The consultant (podiatrist or orthopedist) is paid $50 for the consultation and the Fund pays the costs of any ancillary tests performed by the consultant. Although the program tried to refer half of the individuals to a podiatrist and half to an orthopedist for the second-opinion consultation, it became quite obvious that most orthopedic surgeons refuse to see patients with common podiatric problems such as ingrown toenails, corns, warts and other minor conditions.

It was the Health Fund's policy not to indicate to the second-opinion consultant (podiatrist or orthopedist) the initial diagnosing physician's specialty. The initial diagnosing physician's name is listed on an evaluation form sent to the consultant, but the degree (MD, DO, DPM) is not, therefore, unless the second-opinion consultant happens to know that Dr. X is a podiatrist, the consultant would not be able to tell whether a podiatrist or an orthopedist initially recommended surgery. It is possible that the consultant could have asked the patient directly whether a podiatrist or an orthopedist recommended surgery. Each consultant placed on the second-opinion consultant panel, however, has agreed to provide an unbiased evaluation of the problem and we have no reason to expect that a consultant purposively would reject the opinion of a "competitior."

Findings

Demographic Profile

Of the 363 individuals who had a second opinion for initially recommended foot surgery, 61.4 per cent were females. The age of the sample ranged from 12 years (a covered dependent) to 70 years with a mean age of 47.6 years. There were no statistically significant differences in age between the sexes.

Among the males in the study, 22.9 per cent were porters, 16.4 per cent were doormen, and an equal percentage were handymen, elevator operators or individuals for whom occupational status was unknown, (12.1 respectively). Among the females in the study, 48.9 per cent were cleaning ladies and 23.3 per cent were covered dependents, the overwhelming majority of whom were nonworking housewives. The remaining individuals of both sexes were either building superintendents or building security guards.

One third (30 per cent) of the sample initially were recommended for a bunionectomy (bilateral or unilateral), and 22 per cent initially were recommended for surgical correction of hammertoes, two common orthopedic conditions. Roughly 15 per cent initially were recommended for surgical removal of an ingrown toenail. The remainder were initially recommended for other common podiatric conditions.* Of those initially recommended for common orthopedic conditions, over half (59.9 per cent) had multiple diagnoses; i.e., removal of bunion and correction of hammertoes. Of those initially recommended for common podiatric conditions, 21.2 percent had multiple diagnoses, i.e., removal of corns and warts.

Second-Opinion Consultant's Recommendation

Approximately one-quarter (27.8 per cent) of those initially recommended for podiatric surgery were seen by an orthopedic surgeon (Board Certified) and 72.2 per cent were seen by a podiatrist. Roughly 82 per cent of the study population was confirmed for surgery. There were no statistically significant differences in the confirmation rate between the males and the females or among age groups. An almost identical percentage within each age and sex category were confirmed for surgery.

A closer inspection of the data shows that the confirmation rate among those seeing a podiatrist was 94.3 per cent as compared to a confirmation rate of 49.5 per cent among those who were seen by an orthopedist. Whereas the podiatrists did not confirm the need for surgery in 5.7 per cent of the cases, the orthopedists did not confirm the need for surgery in 50.1 per cent of the cases.

Data presented in Table 1 show that, with the exception of those initially recommended for surgery to correct a heel spur, the overwhelming majority within each podiatric diagnostic category were confirmed for surgery.

When the data were analyzed by the second opinion consultant's specialty, dramatic differences were apparent. Table 2 shows the confirmed/not confirmed status

---

* Corns, metatarsalgia, warts, neuroma.

TABLE 1. Podiatric Conditions Analyzed by Confirmed/Not Confirmed Status

Condition	Confirmed		Not Confirmed		Total	
	N	%	N	%	N	%
Bunion	88	80.0	22	20.0	110	100.0
Corn	24	92.3	2	7.7	26	100.0
Hammertoe	75	86.2	12	13.8	87	100.0
Heel spur	10	43.5	13	56.5	23	100.0
Ingrown toenail	54	100.0	0	0	54	100.0
Neuroma	10	90.9	1	9.1	11	100.0
Osteoarthritis	6	75.0	2	25.0	8	100.0
Other	31	70.1	13	29.5	44	100.0
Total	298	82.1	65	17.9	363	100.0

of the podiatric conditions screened by the podiatrists. For every condition, the podiatrists overwhelmingly confirmed the need for surgery. Conversely, Table 3 shows that the orthopedic second-opinion consultant was less likely to confirm the need for surgery in each category. Whereas the podiatrists confirmed the need for a bunionectomy in 95.8 per cent of the cases, the orthopedists confirmed 50 per cent of the potential bunionectomies. Whereas the podiatrists confirmed the need for surgery in 93.7 per cent of the cases with hammertoes, the orthopedists confirmation rate for hammertoes was 66.7 per cent. Whereas the podiatrists confirmed the need for repair of a heel spur in 75.0 per cent of the cases, only one individual was confirmed for this operation by an orthopedic consultant of the eight cases seen.

Table 4 shows the distribution of consultant reasons for nonconfirmation analyzed by podiatric conditions. In one third of the the cases, the consultants felt that corrective footwear was preferable to surgery and in almost 17 per cent of the cases, the consultants felt that the particular procedure recommended initially was incorrect. Of those not confirmed for a potential bunionectomy, in 27.8 per cent of the cases the second-opinion consultants felt that the initial procedure recommended was incorrect and in 22.2 per cent of the cases corrective footwear was recommended instead of surgery. In 38.5 per cent of the cases, of those not confirmed for

TABLE 2. Confirmed/Not Confirmed Status Analyzed by Podiatric Conditions Screened by a *Podiatrist*

Condition	Confirmed		Not Confirmed		Total	
	N	%	N	%	N	%
Bunion	69	95.8	3	4.2	72	100.0
Corn	17	94.4	1	5.6	18	100.0
Hammertoe	59	93.7	4	6.3	63	100.0
Heel spur	9	75.0	3	25.0	12	100.0
Ingrown toenail	52	100.0	0	0	52	100.0
Neuroma	10	100.0	0	0	10	100.0
Osteoarthritis	6	100.0	0	0	6	100.0
Other	26	89.7	3	10.3	29	100.0
Total	248	94.7	14	5.3	262	100.0

TABLE 3. Confirmed/Not Confirmed Status Analyzed by Podiatric Conditions Screened by an *Orthopedist*

Condition	Confirmed		Not Confirmed		Total	
	N	%	N	%	N	%
Bunion	19	50.0	19	50.0	38	100.0
Corn	7	87.5	1	12.5	8	100.0
Hammertoe	16	66.7	8	33.3	24	100.0
Heel spur	1	9.1	10	90.9	11	100.0
Ingrown toenail	2	100.0	0	0	2	100.0
Neuroma	0	0	1	100.0	1	100.0
Osteoarthritis	0	0	2	100.0	2	100.0
Other	5	33.3	10	66.7	15	100.0
Total	50	49.5	51	50.5	101	100.0

correction of a heel spur, the consultants felt that corrective footwear was preferable. Reasons for nonconfirmation varied more among those recommended initially for correction of hammertoes, but the data should be interpreted with caution due to the small numbers involved.

Differences were apparent in reasons for nonconfirmation by the second-opinion consultant between the orthopedists and the podiatrists. In general, the orthopedists were more conservative in their recommendation for surgery. In 19 out of 20 cases, it was the orthopedist who suggested that corrective footwear was preferable to surgery. In 9 out of 11 cases, the orthopedist felt that the initially recommended procedure was incorrect and that the specifically recommended procedure should not be performed. In four out of six cases, the orthopedist felt the surgical procedure was contraindicated, and in five out of six cases, the orthopedist felt that the symptoms were not severe enough to warrant surgery. As was shown earlier, the podiatrists were more likely to confirm the need for surgery.

The decision to have surgery is made by the patient, regardless of the consultant's recommendation. Of those who were not confirmed, 81.5 per cent did not have surgery. Conversely, of those who were confirmed for surgery, the majority (60.1 per cent) did not have the operation performed. There was no statistical difference in surgical status between the males and females. Data show that 52.9 per cent of those seen by a podiatrist decided to have surgery, whereas 66.3 per cent of those seen by an orthopedist decided against surgery. The majority of those who did have the surgery performed did so within two months from the date of their consultation (76.4 per cent), and 90 per cent had the surgery performed within three months. In 95 per cent of the cases in which surgery was performed, the surgeon was a podiatrist. Of those who had surgery, 52.9 per cent had the procedure performed in an office. Conditions such as corns, ingrown toenails, removal of warts and other minor podiatric conditions were performed on an outpatient rather than on an inpatient basis.

In general, the podiatrists ordered more ancillary tests than did the orthopedists. Podiatrists ordered additional x-rays in 38.3 per cent of the cases, compared to the orthopedists who ordered additional x-rays in 11.5 per cent of the cases. Each patient was instructed to bring x-rays with them to the second-opinion consultation. Since most of the patients were seen initially by one of the Fund's Health Center doctors, we can state with assurity that the x-rays were handed to the patient a day before the second-opinion consultation. Others had to obtain their x-rays from their own private physicians or from the hospital. Each patient is reminded to bring x-rays to the consultation during a telephone conversation by one of the second-opinion referral officers. However, of the 72 potential bunionectomies seen by the podiatrists, 29 (40.4 per cent) had additional x-rays taken. Conversely, of the 38 cases of potential bunionectomies seen by the orthopedists, two (5.3 per cent) needed additional x-rays.

Of the 63 individuals with hammertoes seen by the podiatrists, 32 (50.8 per cent) had additional x-rays taken. Conversely, only two of the 24 individuals with hammertoes seen by the orthopedists needed additional x-rays taken (8.3 per cent). Of the 12 individuals with a heel spur seen by the podiatrists, (N = 8; 66.7 per cent) had additional x-rays taken as compared to two of the 11 individuals with a heel spur seen by the orthopedists (18.2 per cent).

Of 18 individuals with corns seen by the podiatrists, five (27.8 per cent) had x-rays taken as part of the second-opinion examination. None of the eight individuals with corns who were seen by an orthopedists needed additional x-rays. The cost of the additional x-rays taken by the podiatrists total $3,700.00.

Economic Considerations

The surgical bill for the major podiatric conditions performed in hospitals§ totaled $104,071; the mean cost was $1169.00. For those procedures performed in the office (outpatient) the surgeons' bills totaled $31,717; the mean was $488. The average total expense for inhospital podiatric surgery totaled $2,921 while the average total expense for in-office surgery totaled

TABLE 4. Reasons for Nonconfirmation Analyzed by Podiatric Conditions

Condition	Corrective Footwear Preferable		Incorrect Procedure		Symptoms Not Severe Enough		Surgery Contraindicated		Patient Asymptomatic		Medication Preferable		Other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Bunion	4	22.2	5	27.8	2	11.1	3	16.7	2	11.1	0	0	2	11.1	18	100.0
Corn	1	20.0	0	0	2	40.0	0	0	0	0	2	40.0	0	0	5	100.0
Hammertoe	1	16.7	2	33.3	0	0	1	16.7	1	16.7	0	0	1	16.7	6	100.1
Heel spur	5	38.5	0	0	1	7.7	2	15.4	0	0	1	7.7	4	30.8	13	100.1
Neuroma	0	0	1	50.0	0	0	0	0	0	0	1	50.0	0	0	2	100.0
Osteoarthritis	0	0	1	33.3	0	0	0	0	1	33.3	0	0	1	33.3	3	99.9
Other	9	50.0	2	11.1	1	5.6	0	0	0	0	0	0	6	33.3	18	100.0
Total	20	30.8	11	16.9	6	9.2	6	9.2	4	6.2	4	6.2	14	21.5	65	100.0

---

§ Bunions, removal of corns, hammertoes, heel spur.

TABLE 5. Surgical Bills for Specific Podiatric Conditions

Condition	Bill	Average Bill	Range	
			Hi	Lo
Bunion	$ 90199.00	$1610.70	$3091.00	$300.00
Corn	5472.00	390.86	967.00	280.00
Hammertoe	33345.00	741.00	2352.00	165.00
Heel spur	4547.00	757.83	1285.00	410.00
Ingrown toenail	5775.00	160.42	390.00	55.00
Other*	22748.00	733.81	2313.00	125.00
Total	$162086.00	$ 862.16		

* Neuroma, osteoarthritis, other, not specified.

$494. Table 5 shows the costs for the major podiatric conditions. The surgeons' bills for bunionectomies totaled $90,199, with a mean of $1,610.70 and a range of $300 to $3,091. The bills for removing corns averaged $390.86 and for removing of ingrown toenails averaged $160.42. The surgeons' bills for correction of hammertoes totaled $33,345, with a mean of $741.00 and a range of $165 to $2,352. the surgeons' bills to correct heal spurs totaled $4,547.00 with a mean of $757.83 and a range of $410 to $1,285.

One explanation for the large range in bills for specific procedures is that a bilateral bunionectomy was performed or that several toes were corrected at the same time. Also, several procedures were performed at one time but were counted as one procedure (bunionectomy, osteotomy, tendon lengthening and capsulotomy performed at one time). Some individuals also suffered complications, which served to inflate the surgeon's bill.

Since the Fund pays "reasonable and customary" costs, it is likely that the patient would have to pay the difference between what is charged by the surgeon and what is paid for or reimbursed by the Fund. We found that the podiatrists, in particular, were charging in excess of the reasonable and customary costs, but were more than likely to accept the amount paid by the Fund. Specifically, the podiatrists billed a total of $64,870 for in-office surgery, but accepted payments totaling $36,652. For example, the Fund was billed $8,415 for ingrown toenails, but only paid $5,635. The podiatrists billed a total of $240,755 for in-hospital surgery, but accepted $104,071—less than half of what they billed! The average bill was $2,700 and the average reasonable and customary payment was $1,169.

The cost of the consultations totaled $18,100, the reimbursement of the surgeries performed[||] totaled $288,763 and reimbursement of ancillary tests totaled $4,950, for a grand total of $311,813. If a second-opinion consultation is not required for removal of corns, warts and ingrown toenails, the cost of the program would be much less.

Discussion and Conclusion

Until fairly recently, most podiatrists could not obtain hospital privileges and were content to treat the outer tissues. Now, podiatrists are pressing to operate in hospitals and to treat the whole foot (muscles, bones, nerves, tendons and ligaments). Financial motivation certainly plays a part, as does the upgrading of the

---

[||] In and out of hospital.

instruction in podiatry schools. Since major third-party coverage does not reimburse for routine podiatry care, surgery has become the focus of podiatry's push.[6] Operating on the foot brings in more money than trimming and paring. Since most states will not permit podiatrists to administer general anesthesia to office patients, podiatrists must rely on local anesthesia, using minimum incision surgery.

Orthopedists contend that such surgery can cause substantial trauma to surrounding tissues and that the bone being operated upon is not visible. Naturally, they feel that the orthopedist, not the podiatrist, is the most competent to operate on the foot. Orthopedists do have more extensive surgical and medical training and are better able to handle complications.

Even though many foot problems do not require surgery, foot surgery is a very lucrative business. Statistics show that millions of individuals have had their foot problems treated medically and/or surgically, at a cost of millions of dollars, to relieve foot pain or discomfort. The findings from the 32B-J podiatry second-opinion program clearly show the predilection among podiatrists to confirm podiatric surgery and to order an excessive amount of ancillary tests prior to and after surgery. Analysis of the program's first year has made it crystal clear that thousands of dollars are being spent on questionable podiatric surgery. Since the orthopedic consultants refused to examine ingrown toenails and corns, and since almost all second-opinion podiatrists confirmed the need for surgery for these procedures, the Fund decided in April, 1981, to no longer require a second opinion for these procedures.

Regarding the more complicated podiatric conditions (the bunionectomies, the hammertoes and the heel spurs), a second-opinion consultation definitely should be required. It is our opinion that the consultant should be an orthopedic surgeon. These procedures usually are performed in the hospital and have high rates of nonconfirmation. They also are the more costly procedures and substantial savings could be realized if a second-opinion consultation was required. It is also clear that the podiatrists in this study tended to take advantage of a situation and not only were charging excessive fees, but probably were performing many procedures unnecessarily. In April 1981, the Fund also announced that podiatrists were no longer eligible to serve as second-opinion consultants. Only orthopedists are utilized as consultants for all foot-related problems for which a second-opinion consultation is required.

References

1. Greenberg L. An assessment of foot health problems and related health manpower utilization and requirements. J Am Podiatry Assoc 1977;67:102.

2. Gould N, Schneider W, Ashikaga T. Epidemiological survey of foot problems in the continental United States: 1978–1979. Foot and Ankle 1980;1:8.

3. National Center for Health Statistics, Hospital Discharge Survey, Personal communication.

4. Mattera MD. Podiatrists: Pushing for equal rights on the medical team. Medical Economics 1979 (Oct. 15), 159.

5. Fowler EM. Demand likely for podiatrists. New York Times 1980 (Sept. 10), D 19.

6. What to do when your feet hurt. Consumer Reports, March 1981, pp 142-146.

MEDICAL CARE
August 1982, Vol. XX, No. 8

Utilization of an Automated Multiphasic Health Testing System for Performing Prehospitalization Examinations

RAFAEL S. CAREL, M.D., M.P.H.,* ERNESTO KAHAN, M.D., M.P.H., YAAKOV HART, M.D., AND NAOMI PANUSH, M.SC.

Various impacts of performing prehospitalization examinations (of elective surgical patients) by utilizing automated multiphasic screening technology were evaluated. Compared to the conventional way of conducting such examinations (in the concerned health care organization) the automated system was found to be more efficient (results were available earlier, in a more complete and convenient format), saved considerable time to patients and resulted in a shorter average hospital (postoperative) stay, suggesting that the automated technology is also cost effective. Postoperative stay was reduced by 20 per cent (1.8 days) in the group of patients who undertook the automated multiphasic testing, as compared to a similar group of patients who performed the required tests in the conventional way. Possible explanations for this finding are discussed.

DIFFERENT HEALTH CARE organizations have approached and solved the problem of elective prehospitalization testing in different ways. The health care organization under consideration is a large, prepaid, comprehensive health plan which has been described previously.[1] To facilitate and shorten hospital stay prior to the planned operation, each surgical ward has established a routine battery of tests as a prerequisite for admission. This battery includes the following: physical examination; blood pressure; ECG; chest roentgenogram; bleeding, clotting and prothrombin time; basic blood tests (CBC, urea, glucose, electrolytes, etc.) and urinalysis. Specific tests related to evaluation of the major illness such as x-ray series, biopsies or cytological examinations usually are performed in outpatient (specialist) clinics. The routine, basic prehospitalization examinations are conducted under the supervision of the family physician in the regular clinics and laboratories of the concerned health organization. Hospitalization and surgery dates are scheduled well in advance (by the specific surgical service) for most patients. Each ward is allocated fixed days for using operating theaters and scheduling of elective cases is planned accordingly.

This study was to evaluate the currently used method of executing prehospitalization examinations and to compare it with

---

From the Department of Preventive and Social Medicine, Sackler School of Medicine, Tel-Aviv University, Israel.

* Dr. Carel is also from the MOR Institute for Medical Data.

Reprint requests: R. S. Carel, M.D., Department of Preventive and Social Medicine, Sackler School of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel.

0025-7079/82/0800/0871/$01.05 © J. B. Lippincott Co.

performing the same battery of tests by utilizing an automated multiphasic health testing system (AMHTS).

Material and Methods

Figure 1 is a block diagram of the various steps necessary for completing the required tests by both modalities studied—the conventional one (steps 1 to 3) and the automated one (steps 1 and 2). In the conventional "track," the family physician refers the patient to several different laboratories (step 2) to acquire all the necessary medical data included in the prehospitalization profile, while in the AMHTS track all these tests are performed in one facility (step 1) by a dedicated unit.

The study group included 155 subjects. Seventy-eight patients (63 males and 15 females) were tested in the conventional way and 77 (63 males and 14 females) by the AMHTS. Each study group included all cases scheduled for elective surgery in one urological ward during six consecutive months (3 months for each group). The concerned ward draws its patients from a defined geographic area and a uniform mixture of cases is scheduled through the whole year. Age distribution of the subjects is presented in Table 1, showing a similar distribution for both groups. Mean age in the conventional group was 66.1 years and that of the AMHTS group was 63.0 years. Most of the subjects in both groups were over 60 years old (about 70 per cent). Distribution of examinees by types of diagnoses (ICDA definitions) is given in Table 2. About 40 per cent of the subjects in both groups were admitted for prostatectomy due to benign hypertrophy of the prostate and another 10 per cent for malignant tumor of the prostate. Approximately 20 per cent were hospitalized because of other suspected genitourinary malignancies. All other surgical problems constituted less than 30 per cent for both groups. Thus subjects in both groups were quite similar with respect to age and sex distribution, types of diagnoses, severity of illness, residential area and socioeconomic status, as verified by a questionnaire, indicating that selection bias could not be a major factor.

Results

Patients who were examined by the conventional method required, on the average, two work days to complete the necessary tests. Thirty people (39 per cent) completed the examinations in 1 day, 19 (24 per cent) in 2 days and 17 (22 per cent) in 3 days. In 12 subjects, exact information could not be obtained. All the examinees who were tested in the AMHTS completed

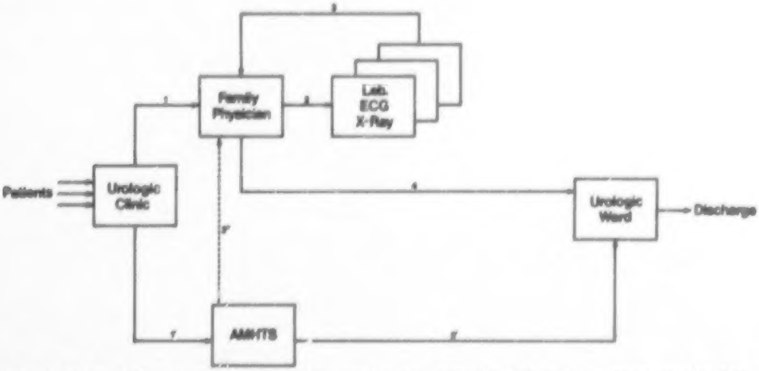


FIG. 1. Flow chart of the steps involved in conducting preadmission examinations by the conventional and AMHTS methods.

the procedure in less than 4 hours (34 in less than 2 hours, 36 in less than 3 hours and 4 in 3–5 hours). The major reason for the long duration of the conventional examinations was that patients had to apply to different laboratories, which often were in different locations. For performing the various tests, 40 patients (51 per cent) visited only one facility, 18 (23 per cent) visisted two facilities and 20 (26 per cent) visited three facilities. For this group, final results of the various tests were received by the family physician (step 3, Fig. 1) within 2 weeks with 74 (94 per cent) patients, 2–3 weeks with two (3 per cent) patients and with another two in over three weeks. On the average it took 8.1 days until the family physician had all the required test results at this disposal. In the AMHTS method all results were collected, collated and printed as a uniform report by a computer within 3 days and transferred directly to the concerned urological ward. It should be realized that hospitalization (and operating) dates were not affected in any way by the method of performing preadmission examinations. Each patient had a fixed admission date independent of how soon the required tests were obtained (or the method of prehospitalization testing). Therefore, even though all patients tested by the AMHTS had their results ready within 72 hours, only 45 (58 per cent) were hospitalized within 2 weeks of their prehospitalization examinations.

No difference between the two groups, with respect to the time interval between admission and operation, could be demonstrated. This could be expected because elective admission date and the day of surgery are predetermined for each surgical service in the concerned hospital (and usually are independent of the prehospitalization tests). Almost all of the examinees in both groups were admitted on their scheduled date.

In Figure 2, the cumulative distribution of the time interval between day of surgery and day of discharge is shown. Fifty per cent of AMHTS subjects were discharged less than 7 days after their operation, while in the other group only 30 per cent of the patients were discharged 7 days after surgery. The average postoperative stay in the AMHTS group was 6.9 days and that in the "conventional" group was 8.7 days (1.8 days longer). The difference between the means of postoperative stay of both groups is statistically significant ($p < 0.001$).

TABLE 1. Distribution of Examinees by Age in Both Study Groups. (Numbers in Parentheses are Percentages of the Same Study Group)

Age Groups (years)	Conventional n	AMHTS n
0–49	8 (10)	9 (12)
50–59	7 (9)	15 (20)
60–69	32 (41)	26 (34)
70 +	31 (40)	27 (34)
Total	78 (100)	77 (100)

Discussion

This paper addresses itself to a specific problem in delivery of health care in a public health organization: that of performing routine prehospitalization examinations for elective surgery in an efficient way. It has felt for a long time that the existing method for conducting these examinations, either during the hospital stay or by referring patients to existing (usually overloaded) laboratories and clinics, is an inefficient method that creates an extra burden on ambulatory services. Not only is it a time-consuming process for those who must be examined, but because of its fragmented nature, it also puts extra demands on physicians and medical staff in performing many routine tasks and much paper work. Moreover, it was found that in many instances the information gathered by these procedures was incomplete, not well organized and of variable quality. Thus,

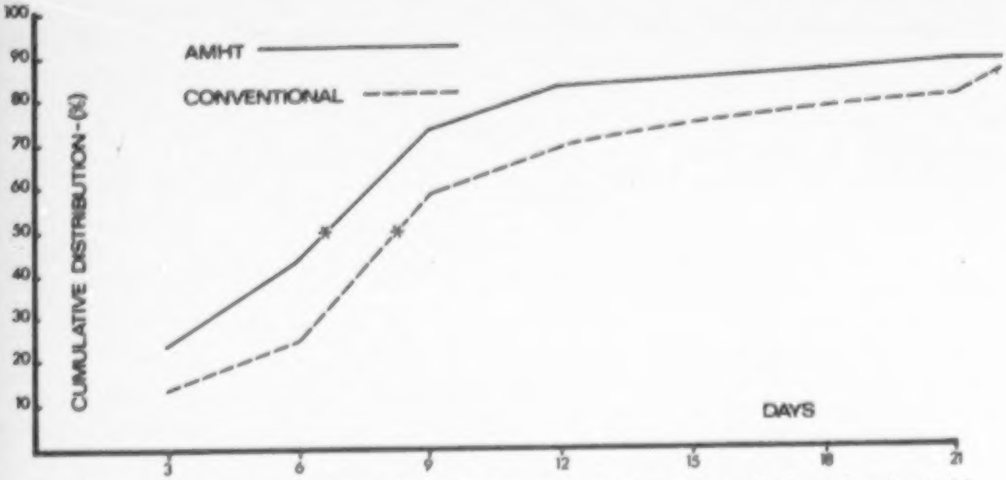


FIG. 2. Cumulative percentage distribution of the time duration from surgery to discharge from hospital for both groups.

new, more efficient (and tentatively more cost-effective) methods were explored. Utilization of multiphasic screening technology as a possible answer to this issue was investigated. A broad range of literature describing applications of AMHTS in compliance to various needs of ambulatory health care services[1–9] and, more specifically, as a part of hospital setups,[10,11] is available. AMHTS are particularly suitable for performing such uniform examinations, as is required for the concerned prehospitalization routine. Multiphasic screening units are dedicated systems geared for rapid and convenient processing of various populations. Most testing procedures are conducted by paramedical personnel and have built-in quality-control procedures. Standard concise reports are generated automatically and rapidly by the computer, and transferred to the appropriate ward. Addition-

TABLE 2. Distribution of Examinees by Diagnoses. (Numbers in Parentheses are Percentages of the Same Study Group)

Diagnoses	Conventional	AMHTS
Benign hypertrophy of prostate	36 (46)	28 (35)
Neoplasms of geniourinary organs	12 (15)	19 (24)
Malignant neoplasms of prostate	5 (6)	10 (13)
Urinary calculi	3 (4)	9 (11)
Other diseases of urinary system	9 (12)	2 (3)
Other diseases of male and female genital organs	6 (8)	5 (6)
Other ill-defined conditions	7 (9)	6 (8)
Total	78 (100)	78 (100)

ally, the same report also is sent to the personal physician (step, Fig. 1).

By referring patients for prehospitalization examinations to an AMHTS, several benefits could be demonstrated. The individuals who were examined saved a great deal of time (less than 4 hours vs. 2.1 days). Primarily, this is due to the fact that in the AMHTS track all tests were performed in one facility and controlled order (Fig. 1). Physicians were supplied with a printed, organized report, not fragmentary information that had been accumulated over a period of several weeks. Laboratories and staff of the health care organization were relieved of the time-consuming routine tasks involved in these examinations.

The finding that patients who were tested by the AMHTS had shorter postoperative stays in the hospital (by 1.8 days on the average) was not anticipated. Both study groups were quite similar in their age-sex-disease characteristics (Tables 1 and 2).

Similarly, no difference between the groups with respect to admission or operative diagnoses was found. Thus, none of these factors seemed to be responsible for the difference in the postoperative length of stay related to a possible bias affecting this outcome. Since there was no reason to assume that a postoperative stay would be affected by such basic and routine procedure, no specific questions in the evaluation questionnaire addressed themselves to this issue. This important finding (especially when substantial savings in hospitalizations costs are implied) requires further evaluation. By the same token, the feasibility of shortening preoperative hospital stay for AMHTS patients is now under investigation. This later issue seems to be primarily an administrative problem, requiring the introduction of some changes in scheduling admission and preoperative routines.

Results of this study support previous claims[1] that AMHT technology could be an efficient modality in solving particular needs in various segments of the health care system. Identifying such processes, where AMHT can be shown to provide more rapid, convenient or costeffective service (compared to the existing one) could lead to improvements in the delivery of care in many areas. Exploring new potential utilizations of AMHTS by different health care organizations also could assist in developing more versatile and extensive uses of this technology, therefore improving the overall efficiency and quality of the concerned health services.

References

1. Carel RS, Leshem G. Evaluation of the cost effectiveness of an automated multiphasic health testing system. Prev Med 1980;9:689.

2. Collen MF, ed. Multiphasic Health Testing Services. New York: John Wiley and Sons, 1978.

3. Dales LG, Friedman GD, Collen MF. Evaluating periodic multiphasic health checkups: a controlled trial. J Chronic Dis 1979;32:385.

4. Gilbert F, Nordyke R. Automated multiphasic health testing in multidisciplinary group practice. Prev Med 1973;1:261.

5. Fletcher SW, Sourkes M, Rabzel J, et al. Multiphasic screening, case finding tool in a teaching hospital medical clinic. JAMA 1977;237:887.

6. Holocomb FW. IBM's health screening program and medical data system. Journal of Occupational Medicine 1973;11:863.

7. Hsieh RKC. Multiphasic health testing in a public health service hospital. Med Instrum 1973;8:285.

8. Shapiro S, Frink R, Rosenberg C. A program to measure the impact of multiphasic health testing on health differentials between poverty and nonpoverty groups. Med Care 1972;10:207.

9. Yedidia A. California cannery workers program. Arch Environ Health 1973;27:259.

10. Davey PW, Letts HW, Anderson PA. Experience with a hospital laboratory admission screen. Can Med Assoc J 1970;103:140.

11. Durbridge TC, Edwards F, Edwards RB. An evaluation of multiphasic screening on admission to hospital. Med J Aust 1976;1:703.

BEST AVAILABLE COPY

Medical Care

September 1982
Vol. 20 No. 9
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Barriers to Medical Care of Mexican-Americans: The Role of Social Class, Acculturation, and Social Isolation

Racial/Ethnic Differences in Access to Health Care: Further Comments on the Use–Disability Ratio

Does Income Affect Mortality? An Analysis of the Effects of Different Types of Income on Age/Sex/Race-Specific Mortality Rates in the United States

Income and Race Differences in Children's Health in the Mid-1960s

Regional Differences in Hospital Use Standardized by Reported Morbidity

Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population?

Regression Analysis in Health Services Research: The Use of Dummy Variables

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL
BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Marion Athearn
Jennifer Gardon
Editorial Assistants

S. E. BERKI, M.A.
University of Michigan

ROBERT H. BROOK, M.D., SC.D.
University of California
Los Angeles

GORDON H. DE FRIESE, PH.D.
University of North Carolina

EUGENE FEINGOLD, PH.D.
University of Michigan

BRANT E. FRIES, PH.D.
Yale University

THOMAS R. GODKINS, B.S., P.A.
University of Oklahoma

HAROLD HUNTER, DR. P.H.
Washington, D.C.

SHERMAN A. JAMES, PH.D.
University of North Carolina

ANTHONY L. KOMAROFF, M.D.
Brigham and Women's
Hospital, Boston

JAMES P. LOGERFO, M.D., M.P.H.
University of Washington

NANCY MILIO, PH.D.
University of North Carolina

DUNCAN NEUHAUSER, PH.D.
Case Western Reserve University

JOHN E. WARE, JR., PH.D.
The Rand Corporation

ALONZO S. YERBY, M.D.
Harvard University

ALFRED YANKAUER, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year for individuals, $48.00, for institutions, $55.00; special price U.S. students, $35.00; all other countries except Japan, one year, for individuals, $59.00, for institutions, $68.00. Single copies $6.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, MD, and additional mailing offices.

September 1982
Volume 20, Number 9

CONTENTS

POLICY FORUM

Inequity in Access to Health Care: An Old and Persistent Worry. *Miriam Orleans* 881

ORIGINAL ARTICLES

Barriers to Medical Care of Mexican-Americans: The Role of Social Class, Acculturation, and Social Isolation. *Alan P. Chesney, Juan A. Chavira, Rogers P. Hall, and Howard E. Gary, Jr.* 883

Racial/Ethnic Differences in Access to Health Care: Further Comments on the Use–Disability Ratio. *Alfred C. Marcus and Jeffrey D. Stone* 892

Does Income Affect Mortality? An Analysis of the Effects of Different Types of Income on Age/Sex/Race-Specific Mortality Rates in the United States. *Jack Hadley and Anthony Osei* 901

Income and Race Differences in Children's Health in the Mid-1960s. *Linda N. Edwards and Michael Grossman* 915

Regional Differences in Hospital Use Standardized by Reported Morbidity. *Mark S. Blumberg* 931

Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population? *Noralou P. Roos and Leslie L. Roos, Jr.* 945

Regression Analysis in Health Services Research: The Use of Dummy Variables. *Lincoln Polissar and Paula Diehr* 959

COMMUNICATION

Assessing Needs for Community Hospital Continuing Medical Education. *Charles E. Osborne* 967

PUBLICATIONS RECEIVED 972

MEDICAL CARE
September 1982, Vol. XX, No. 9

Original Articles

Barriers to Medical Care of Mexican-Americans: The Role of Social Class, Acculturation, and Social Isolation

ALAN P. CHESNEY, PH.D.,* JUAN A. CHAVIRA, PH.D.,† ROGERS P. HALL, M.A.,‡ AND HOWARD E. GARY, JR., PH.D.§

Literature on medical care utilization by Mexican-Americans suggests that patterns of utilization can be explained, in part, by acculturation, social class and social isolation. The relative importance of these variables is explored in a survey of a rural Mexican-American community in Southern Texas. Data on 152 families are analyzed using a logit analysis, with utilization as the outcome variable. The results suggest that acculturation has a direct effect, while social class and social isolation interact to have a weak effect independent of acculturation. These results confirm the importance of understanding all three variables in planning and implementing medical care programs in Mexican-American communities.

RESEARCH on Mexican-American health care utilization may be categorized by the conceptual framework that guides the research. From an anthropologic framework, health behaviors are explained by referring to characteristics of the culture such as folk beliefs, acculturation, traditionalism and attitudes. Research conducted by Saunders,[1] Clark,[2] Rubel,[3] Welch,[4] and Farge[5] tend to support this explanation of health care utilization.

From a sociologic framework, health behaviors are explained in terms of the relationship of an individual or a family to the community. Such concepts as social isolation, social participation and social class have been related to utilization of health services. Research by Nall and Spielberg,[6] Bullough[7] and Hoppe and Heller[8] has supported the importance of sociologic factors in understanding health behaviors of Mexican Americans.

In some research, both conceptual frameworks have been used. Since traditionalism tends to be associated with isolation from the community, Reeder and Berkanovic[9] used these variables together to explain American health behavior. Social and cultural variables also have been combined by including both in the definition of a major study variable, such as social class.[10] None of these studies however, has analyzed the relative importance of cultural and social factors in studying health care utilization at a community level. Frequently, the cultural and social variables are linked so closely that it is difficult to separate them. Our study is directed to-

---

* Department of Psychiatry and Behavioral Sciences University of Texas Medical Branch, Galveston.

† School of Law, University of Texas, Austin.

‡ Department of Psychiatry and Behavioral Sciences University of Texas Medical Branch, Galveston.

§ University of Houston, Clear Lake City.

Data reported in this paper were collected pursuant to contract #232-78-0010 from the Bureau of Health Manpower, Department of Health, Education and Welfare.

Reprint requests: Alan P. Chesney, Department of Psychiatry & Behavioral Sciences, UTMB, Galveston, TX 77550.

0025-7079/82/0900/0883/$01.25 © J. B. Lippincott Co.

ward this problem. We attempt to identify the relative importance of social class, acculturation and social isolation in explaining health care utilization in a rural Mexican-American community.

Sociologists who have studied the relationship between minority groups and the dominant society stress the importance of the conceptual distinction between social class and culture.[11,12] Culture refers to norms and values which serve as guideposts or standards for behavior. Social structure, on the other hand, refers to the set of crystallized social relationships the members of society have with each other and with the society. This study preserves the conceptual distinction between cultural variables (cultural assimilation) and social structure variables (social class and social isolation). Such a distinction will enable us to explore the relative contributions of these variables to medical care utilization by Mexican-Americans.

Setting

The current study was conducted in Hidalgo, a small town in the Lower Rio Grande Valley of Texas. Hidalgo is located on the banks of the Rio Grande, 40 miles from the mouth of the river. The 2,500 residents of Hidalgo are predominantly Mexican-American (85 per cent) and most are poor. Approximately 60 per cent of the city's families live on less than $4,000 per year.

Hidalgo is connected to other towns in the Valley by Route 281, which ends at the International Bridge connecting Hidalgo with Reynosa, a large urban area with over 200,000 inhabitants. As a gateway to Mexico, Hidalgo serves as a point of entry into Mexico for nine million people per year. Since Hidalgo shares the revenue from the bridge, the town is in the unique position of public prosperity, despite the poverty of its residents. For example, the city has a modern police and fire station, it has a new high school building and all of its streets are paved.

Housing in Hidalgo is a study in contrasts: There are numerous houses built of plywood and cardboard that have dirt floors, there are mobile homes without wheels and there is an elegant mansion. An enumeration of dwelling units revealed that Hidalgo has 443 dwelling units, 53 per cent of which are in sound condition, 18 per cent are deteriorating and 28 per cent are dilapidated. The criteria for classifying the condition of a dwelling were developed by the United States Public Health Services in southern Texas.[13] The houses frequently are intermixed with sound houses. A few of the houses had unsanitary refuse storage (6 per cent), some had outhouses that could be observed from the road (13.5 per cent) and a small number of houses had livestock living in or near them.

At the time of this study, health services in Hidalgo were limited. There were no physicians or dentists practicing or living in Hidalgo. There were two lay midwives (Parteras) and a school nurse. Ambulance service to a Red Cross Clinic was available to Hidalgo residents. McAllen General Hospital, the nearest United States hospital, was 10 miles away. The Hidalgo County Health Department provided bimonthly clinics in Hidalgo that were mostly preventive. Consequently, many health problems remain undiagnosed and untreated.

Methods

Fifty per cent of the dwelling units were selected at random from the enumeration data. Of the 222 dwelling units included in the sample, 36 were vacant and 16 could not be located. The vacant dwelling units may have been migrant housing; the 16 units that could not be located may have been moved during the time between the enumeration and the survey. From the remaining 170 units, 152 complete interviews were obtained.

Data collection took place during the summer months of 1976. Each family was contacted by an interviewer who explained the purpose of the visit and requested cooperation. Interviews were conducted in English or Spanish, according to the respondent's preference. Only adult respondents were interviewed and, whenever possible, this person was the head of the household (either parent).

Social class of the family was operationally defined by family income. (In those cases where family income was unknown, social class was assigned on the basis of occupation.) If both income and occupation are unknown, then education was used to determine social class. These criteria allowed us to categorize each family as being in the upper or lower social class. The use of only two classes was considered appropriate since the community has a small middle class.

TABLE 1. Socio-demographic Characteristics of the 152 Mexican-American Families Interviewed

	N	%
Respondent characteristics		
Sex		
Male	15	9.8
Female	137	90.2
Age		
Under 25	18	11.8
25–34	41	27.0
35–44	23	15.1
45–54	39	25.7
55 and older	31	20.4
Years of formal education		
No formal education	17	11.2
Some, but less than 7 years	74	48.6
7–11 years	34	22.4
Completed high school	21	13.8
Education beyond high school	6	4.0
Family characteristics		
Number of persons in dwelling		
1–2	23	15.1
3–5	84	55.3
6–7	31	20.4
8 or more	14	9.2
Occupation of breadwinner		
Managerial–technical	11	7.2
Sales–clerical	11	7.2
Craftman–skilled labor	15	9.9
Semiskilled labor	22	14.5
Laborers and service workers	36	18.6
Unemployed, never worked, retired	57	37.5
Monthly family income		
Less than $86	21	13.8
$87–$250	29	19.1
$251–$499	38	25.0
$500–$999	31	20.4
$1,000 or more	15	9.9
Unknown	18	11.8

Social isolation was defined conceptually as the opportunity for social interaction, participation and support. This definition is consistent with definitions that have defined social isolation as one component of alienation.[7-14] In addition, this definition allowed us to examine social class and social isolation as separate variables. Previous measures of the social isolation component of alienation rely on the respondent's perception of his social network, e.g., "I can always count on my family for help when I need it."[13]

Meyers,[14] however, has suggested that the conceptual opposite of social isolation, social integration, may be measured more precisely by focusing on demographic and behavioral indicators. Therefore, the following demographic and behavioral data were combined into an index of social isolation: (1) marital status, (2) presence of a telephone in the dwelling unit, (3) amount of travel to Mexico, (4) amount of travel to other Valley towns and (5) length of time residing in Hidalgo.

Acculturation or cultural assimilation is conceptually defined in this research as "change of cultural patterns to those of the host society."[12] In this research, the term refers to acceptance of Anglo norms and values as the standard for judging one's behavior. Research on Mexican-Americans has documented the importance of language as an essential component of cultural assimilation. Olmedo[16] concluded that the major dimension of acculturation is language proficiency and preference. Olmedo's research also highlights place of birth as a major indicator of acculturation. It is used frequently as a measure of acculturation on the assumption that it reflects location of early socialization. However, since babies may be born in southern Texas and raised in Mexico, location of education has been included with location of birth and language proficiency to form an index of acculturation. The index indicates past value socialization and the present integration of the family into the norms and values of American (Anglo) society.

Data on utilization of health services were collected using two different methods. Respondents were asked the question, "If you or a member of your family is sick, what reasons would prevent you from seeing a doctor?" In addition, respondents were asked if a family member had experienced any of 20 specific symptoms in the past three months and if a physician was consulted if they had experienced them. The list of symptoms is an adaptation of Koos'[17] list. Two major problems emerged in using this list. One is that 27.6 per cent of the respondents reported no symptoms in their family and, therefore, would not be included in the analysis if symptom data were used to determine utilization behavior. The second major problem is that respondents showed a marked (75 per cent) tendency to report only their own symptoms rather than those of their family. An additional problem with the symptom data is that there is no measure of symptom severity. Therefore, the dependent variable is a predisposition to seek medical care given that a perceived need exists. The predisposition is measured by whether the respondent identifies barriers to medical care. This study does not attempt to define variables that are related to the definition of a problem as an illness.

Results

The mean size of the households in the survey was between four and five people. Small households of one or two persons generally were made up of older, retired persons. Single parent households, those in which the head of the household was separated, divorced or widowed, made up 19 per cent of the sample. The mean monthly income per household was approximately $420, 22 per cent of the households reporting either no income or refusing to provide the information. Employment information was collected on the head of the household. Thirty-nine per cent held skilled or semiskilled jobs, 24 per cent held unskilled jobs and 37 per cent were unemployed or retired. The average length of time at the present job for those currently working was 4.2 years. The cost of the housing computed from monthly rent or mortgage payments averaged $74 per month. Houses ranged in size from one to 12 rooms, with an average of approximately five rooms. Thirty-nine per cent of the housing would be classified as crowded by United States Census Bureau of Standards (1.1 or more persons per room). Over 90 per cent of the houses had stoves, refrigerators and radios, but only about half of the homes had running water or telephones.

The survey data on Hidalgo indicate that the sample selected is similar to the community as a whole. In addition, the sample appears representative of other communities along the Rio Grande. Comparison of our demographic descriptions with 1970 census demographic descriptions of the county provides striking similarities with regard to housing conditions, educational levels, income levels and length of residency at present location.

TABLE 2. Measures of the Level Acculturation for 152 Mexican-American Families

	N	%
Location of education		
Mexico	74	48.7
U.S.	78	51.3
Language spoken		
Spanish only	81	53.3
Spanish and English	71	46.7
Location of birth		
Mexico	70	46.1
U.S.	82	53.9
Summed acculturation score		
0,1 low	79	52.0
2,3 high	73	48.0

The first step in analyzing the data from this study is to develop indexes of social class, social isolation and acculturation. The second step is to identify the relationships among the independent variables. The third step is to identify the relative importance of the independent variables in explaining variations in the dependent

TABLE 3. Measures of Levels of Social Isolation for 152 Mexican-American Families

	N	%
Marital status		
Single, widowed, divorced	39	25.7
Married	113	74.3
Telephone in the dwelling unit		
No	76	50
Yes	76	50
Travel to Valley weekly		
No	67	44.1
Yes	85	55.9
Travel to Mexico weekly		
No	95	62.5
Yes	57	37.5
Length of time resided in Hidalgo		
Less than 5 years	67	44.1
More than 5 years	85	55.9
Summed social isolation		
0–3 high	63	41.4
3–5 low	89	58.6

TABLE 4. Distribution of Respondents (N = 152) Predisposition Toward Health Care Utilization by Social Class, Acculturation and Social Isolation

Social Class	Acculturation	Social Isolation	Utilization No	Utilization Yes	Odds of Utilization
Low	Low	Low	15	14	0.93
Low	Low	High	11	6	0.55
Low	High	Low	7	6	0.86
Low	High	High	9	7	0.77
High	Low	Low	10	2	0.20
High	Low	High	13	8	0.62
High	High	Low	5	4	0.80
High	High	High	11	24	2.18

variable, regardless of the association among independent variables.

The measures of acculturation are reported in Table 2. The sample is evenly divided on each of the measures of acculturation and on the overall acculturation score. With regard to social isolation, it appears that the sample tends not to be isolated. Family members reported a tendency to travel to other towns in the valley and to have lived in Hidalgo for more than 5 years. These factors, in combination with a 75 per cent married rate, tend to push the social isolation scores toward low isolation.

In response to the question, "If you or a member of your family were sick, what reasons would prevent you from seeing a doctor?", 53.9 per cent (81) stated that lack of money, transportation or both would stop them from seeing a doctor when they thought it necessary. In contrast, 26.1 per cent (71) reported that they would see a doctor when necessary. Those who identified barriers to physician utilization may be people who would delay in seeking medical care or people for whom medical care is not accessible.

To assess the effects of acculturation, social class and socialization on health service utilization, a logit analysis was performed using a log-linear model.[18–20]

This analysis permits the estimation of how each of the independent variables and their interactions affect the logit (i.e., the log of the odds) of utilization of health services. The contingency table for this analysis is presented in Table 4 and the chi-squared summary table is presented in Table 5. The results of this analysis indicate that only the main effect for acculturation is significant at 0.05 level.

From Table 4 we see that the odds of utilization for the low acculturation group are 0.61 while the odds for the high acculturation group are 1.28. Thus, increasing the level of acculturation from low to high

TABLE 5. Chi-Square Summary Table for the Logits Analysis

Effects	df	χ^{2*}	p
Social class	1	0.01	0.94
Acculturation	1	4.11	0.04
Isolation	1	0.38	0.53
Social class × acculturation	1	2.34	0.29
Social class × isolation	1	3.41	0.06
Acculturation × isolation	1	0.07	0.78
Social class × acculturation × isolation	1	0.13	0.94

* χ^2 = likelihood ratio.

approximately doubles the odds of health service utilization.

Since some researchers may be more comfortable with percentages than with odds, we have converted the estimated odds into percentages using the formula PERCENTAGE = 100 × (ODDS)/(ODDS + 1). Applying this formula to the estimated odds presented above, we find that 38 per cent of the low acculturation group would use health services, while 56 per cent of the high acculturation group would use these services.

A second interesting finding from these data is the suggestion that the effect of isolation on the odds of health service utilization depends on the person's social class. This is tested by the *social class × isolation* interaction in Table 5. This test fails to reach significance at the 0.05 level ($p = 0.07$), but we felt that the finding was sufficiently interesting to present here. Using the procedures outlined above, we estimated the odds of health service utilization separately for each combination of social class and social isolation. We found that for lower social class individuals, the estimated odds of utilization are 1.01 for the low isolation group (i.e., 50 per cent and 0.5 for the high isolation group (i.e., 33 per cent). For upper social class individuals, the odds are 0.788 for the low isolation group (44 per cent) and .854 for the high isolation group (46 per cent). Thus, for lower social class individuals, reducing isolation doubles the odds of health service utilization, while for upper class individuals, isolation essentially has no effect.

Discussion

The results of this community survey suggest two tentative conclusions in explaining medical care utilization by Mexican-Americans. One is that the acculturation is related directly to utilization. Those families that evidence more acculturation are more likely to utilize medical care. The second conclusion is that the impact of social isolation upon Mexican-American utilization is conditional upon social class and independent of acculturation.

This finding resolves the apparent conflict in the literature between a cultural and a social model explaining medical care utilization. In fact, while both cultural and social variables are necessary to explain community patterns of utilization, neither is sufficient to explain the variations in utilization that are described in Hidalgo. The results suggest a new direction for the health sciences research, which focuses on the relative contribution of social and cultural variables in explaining medical care utilization. In addition, this research suggests that analyzing variations in utilization patterns within one ethnic group will be a fruitful avenue of future research.

The conclusions suggested above are tentative because of the limitations of the research. It must be recognized that this survey was conducted in a Mexican-American community where medical care is relatively unavailable and inaccessible. In addition, the community offered what may be unique variations in acculturation and social isolation. Its proximity to Mexico and its population that has a high percentage of Mexican-Americans may have resulted in more families being identified as unacculturated and not socially isolated than would be true for Mexican-Americans nationally.

In addition to the question of generalizability, there is a problem in the precision of measurement. It is desirable to measure utilization of medical care with a prospective study of actual behavior. Use of hypothetical situations leaves a nagging question about its relevance to actual performance. In addition, more sophisticated measures of social class, social isolation and acculturation are in the process of being developed. The measures used in this research are consistent with other efforts to measure these sociologic concepts with demographic variables. More precise measurement will provide a more accurate test of the hypotheses proposed. In developing better instruments, it is important to remember that long structured interviews may not be the most useful method of gathering data about low income ethnic populations. Our experience suggests that a more conversational interview may be more effective.

Keefe and Casas[21] conclude that "Mexican-Americans who are Spanish-speaking, undereducated, and recent residents are for the most part unfamiliar with mental health services." The results of the study suggest that the statement also may be true for health services generally. However, our results suggest that within ethnic groups, variations regarding utilization are not simply a result of lack of knowledge about health care services. One explanation of this finding is that the combination of high income and a social network provide a family with alternatives to medical care. These alternatives may exist within the extended family. High income and high social isolation tend to increase the odds of utilization on the proportion of utilizers is not statistically significant, these trends suggest that higher class families who are socially isolated may lack alternative sources of care and, since they can afford medical care, they will seek it. Interventions which are aimed at increasing medical care utilization of low income, Mexican-Americans shoud focus on acculturation issues such as increasing their awareness of the rules and norms, which are an integral part of medical care. A strategy of reducing social isolation among low income Mexican-Americans as a means to increase medical care utilization is not aimed at the appropriate target group.

The relationship among social class, social isolation and medical care utilization is independent of acculturation. The conditional probability calculations on acculturation and utilization indicate that low acculturation is associated with low utilization. These results suggest that educational programs without regard to social class and social isolation may be beneficial. Since one aspect of acculturation is language, these results suggest that increasing the number of Spanish-speaking medical personnel could result in increased medical care utilization by Mexican-Americans.

These results suggest that the tendency to utilize or not utilize medical care involves a complex social and cultural process. Intervention strategies which oversimplify the process of seeking health care by focusing on either social class, social isolation or acculturation will not produce dramatic changes in utilization rates. We have explored the interaction between these variables in order to document the importance of various population subgroups within the Mexican-American population. Two subgroups tend toward more utilization of medical services. Each group will require unique intervention strategies aimed at increasing medical care utilization.

Future research in this area should take advantage of recent conceptual and methodologic developments in utilization behavior and acculturation. The symptoms–response ratio developed by Anderson and Smedbey[22] offer a more sophisticated measure of underutilization of medical care services. That technique compares a ratio of visits per symptom to the expected ratio of visits per symptom as defined by a panel of physicians. This measure has at least two advantages. First, it eliminates the problem of symptom severity from the equation. Second, it focuses the research on group rates rather than individual behaviors.

Subsequent to our field work, an acculturation scale for Mexican-Americans has been developed by Cueller.[23] The use of an acculturation scale coupled with demographic characteristics (place of birth, place of education, etc.) should provide a powerful measure of acculturation.

The relationship among acculturation, social class, social isolation and medical care utilization proposed here have implications for delivery of health services to rural Mexican-American communities. First, health planners must recognize the diversity that exists among Mexican-Americans. It is not sufficient to develop programs that are based on simplistic notions about "the Mexican-American" population. Some portions of a Mexican-American community find medical care readily accessible. Other segments find the health care system to be unaccessible for very different reasons. Programs geared to increase accessibility of services through education interventions will probably be effective among the segment of the community that is relatively poorly acculturated. However, educational programs are not sufficient; programs which attempt to provide social support mechanisms to increase accessibility will need to be developed carefully. In some cases, the interventionist should link into an existing support system; in other groups a social support system needs to be created. Determining what type of intervention should be implemented to increase health care accessibility in a community requires knowledge of the levels of acculturation, social class and social isolation in the community. Future needs assessments of health care in Mexican-American communities should consider these variables.

Acknowledgment

The authors extend their appreciation to Dr. W. W. Schlottstaedt and the anonymous reviewers for their contributions to this manuscript.

References

1. Saunders L. Cultural differences and medical care. New York: The Russel Sage Foundation, 1954.

2. Clark M. Health in the Mexican-American culture. Los Angeles: University of California Press, 1959.

3. Rubel A. Across the tracks: Mexican-Americans in Texas City. Austin: University of Texas Press, 1966.

4. Welch S, Comer T, Steinman M. Social and attitudinal correlates of health care among Mexican-Americans. J Health Soc Behav 1973;14:212.

5. Farge EJ. La Vida Chicana: Health care attitudes and behaviors of Houston Chicanos. San Francisco: Rand E. Research Associates, 1975.

6. Nall FC, Spielberg T. Social and cultural factors in the responses of Mexican-Americans to medical treatment. J Health Soc Behav 1967;9:299.

7. Bullough B. Poverty, ethnic identity, and preventive health care. J Health Soc Behav 1972;13:347.

8. Hoppe SK, Heller PL. Alienation, familism, and the utilization of health services by Mexican-Americans. J Health Soc Behav 1975;16:304.

9. Reeder L, Berkanovic E. Sociological concommitants of health orientations: A partial replication of Suchman. J Health Soc Behav 1973;14:134.

10. Madsen W. The Mexican-Americans of South Texas. New York: Rinehart & Winston, 1964.

11. Schermerhorn RA. Comparative ethnic relations: a framework for theory and research. New York: Random House, 1970.

12. Gordon M. Human nature, class, and ethnicity. New York: Oxford University Press, 1978.

13. Waldrop RH. Socioeconomic stratification in community health. Washington, DC: U.S. Department of Health, Education, and Welfare 1971:1.

14. Dean D. Alienation: Its meaning and measurement. Am Sociol Rev 1961;24:753.

15. Meyers J, Lindenthal J, Pepper M. Life events, social integration and psychiatric symptomatology. J Health Soc Behav 1975;16:421.

16. Olmedo EL, Padilla AM. Empirical and construct validation of a measure of acculturation for Mexican-Americans. Journal of Soc Psycho 1978;105:179.

17. Koos EL. The health of Regionville. New York: Columbia University Press, 1954.

18. Bishop YMM. Full contingency tables, logits, and split contingency tables. Biometrics 1969;25:383.

19. Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis. Cambridge, Mass.: The MIT Press, 1975.

20. Swafford P. Three parametric techniques for contingency table analysis: a non-technical commentary. Am Sociol Rev 1980;45:664.

21. Keefe S, Casas J. Mexican-Americans and mental health: a selected review and recommendations for mental health delivery. Am J Community Psychol 1980;8:303.

22. Anderson R, Smedbey B. Changes in response to symptoms of illness in the United States and Sweden. Inquiry 1975;12:116.

23. Cueller I. The underutilization of mental health facilities and services by Mexican-Americans: who, where, how much, and why? Proceedings of the Texas-New Mexico Symposium on the Delivery of Mental Health Services to Mexican-Americans, 1979.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Racial/Ethnic Differences in Access to Health Care: Further Comments on the Use–Disability Ratio

ALFRED C. MARCUS, PH.D.,* AND JEFFREY D. STONE, PH.D.†

Indicators of access to health care generally fall into one of two broad groups. The first group of indicators focuses on factors responsible for differential rates of use and often are assessed through reports of patient satisfaction. The second group of indicators focuses more specifically on actual rates of use, and therefore represents a more objective measure of access. One of the most widely used objective indicators of access is the use–disability ratio. This article illustrates some problems in using this ratio in health services research and some strategies one might adopt to minimize these problems. The data reported below were obtained from an analysis of racial/ethnic differences in use of services that was conducted as part of the 1979 Los Angeles Health Survey (N = 1003). Our findings suggest that 1) blacks had greater access to health care than either white/Anglos or Hispanics, particularly for respiratory and musculoskeletal problems, and 2) the use–disability ratio should be used cautiously, since it is sensitive to subgroup differences in the chronicity and cause of the disability.

PERHAPS no concept is as central to health services research as patient access to health care. Ironically, few concepts also are as ill-defined.[1] In an attempt to rectify this problem, Penchansky and Thomas[2] recently introduced a taxonomy of access that focuses on factors hypothesized to influence use of services. According to their model, access is "a general concept that summarizes a set of more specific dimensions describing the fit between the patient and the health care system." The specific dimensions of access identified by Penchansky and Thomas included availability, accessibility, accommodation, affordability and acceptability.

Although the degree of fit between the individual and the health care system, as determined by reports of patient satisfaction, is important to understanding health services utilization, there is also a need to develop objective measures of access that focus on the frequency of using health services.[3] For example, recent data from the National Center for Health Statistics show that people with lower incomes have more contact with doctors per year than people with higher incomes.[4] However, this difference does not mean that low income groups have greater access to health care than middle and upper income groups, since it might also reflect variations in the *need* for health care. As both Aday[4,5] and

---

* Director of Program Evaluation, UCLA Jonsson Comprehensive Cancer Center, Division of Cancer Control, University of California, Los Angeles.

† Research Associate, School of Public Health, University of California, Los Angeles.

Data reported in this paper were collected pursuant to grant number 5-R18-CA-18451 "Processes in Health Behavior and Cancer Control," awarded by the National Cancer Institute to the late Leo G. Reeder, Principal Investigator.

Requests for reprints: Alfred C. Marcus, Ph.D., Director of Program Evaluation, UCLA Jonsson Comprehensive Cancer Center, Division of Cancer Control, 10920 Wilshire Boulevard, Suite 1106, Los Angeles, California 90024.

0025-7079/82/0900/0892/$01.25 © J. B. Lippincott Co.

Andersen[5] observed, access to health care is best considered within the context of whether people who actually need medical care receive it. It is in defining and assessing this need that health status indicators become essential to the measurement of access.

One of the most widely used objective indicators of access is the use–disability ratio, which is defined as the number of doctor visits per 100 disability days. Thus, this ratio combines the features of an epidemiologic rate with a measure of use that has been standardized on the basis of need (i.e., disability days). This ratio made its first appearance in an article by Aday,[4] and since then has been reported in several other national, state and community surveys.[6–10]

Although the use–disability ratio has been used extensively in several prominent studies, it has also had its share of critics. For example, Fox and Bice[11] criticized the index because it assumes a linear relationship between need for care and disability days. Chen[3] has raised two other objections. First, the index has no objective norm or standard, making it difficult to interpret from a policy perspective (e.g., does a higher ratio indicate more appropriate use of health services or overutilization of health services?). Second, Chen notes that the cause of the disability is not taken into account in this index. Consequently, in comparing rates among population subgroups, one must assume that a comparable number of disability days represents a comparable need for health care—regardless of the cause of the disability. Montiero[8] made a similar argumeant in her report on the Rhode Island Health Survey. She criticized the ratio because it failed to distinguish between acute and chronic illness. That is, if the source of the disability is an acute illness, then a visit to the physician may be more likely than if the illness was long-standing or chronic. Accordingly, "since economic levels and age groups are likely to differ in the underlying source of restricted activity days, then differences in the use/need ratios for these groups may to some extent reflect the type of illness experienced rather than any real difference in access to medical care."

In this article, we report findings from a comparison made between the use–disability ratio and another indicator of access. Our purpose is to illustrate some of the problems that may be encountered in using this ratio in health services research and some of the strategies one might adopt to minimize these problems.

Methods and Procedures

The data reported were collected during the summer of 1979 as part of the Los Angeles Health Survey (LAHS). Fieldwork was conducted by the UCLA Survey Research Center of the Institute for Social Science Research (ISSR). ISSR maintains a sampling frame containing approximately 20,000 computer-readable addresses sampled from Los Angeles County on an area probability basis. Samples drawn from the frame may be characterized as three stage samples with probabilities proportional to size.[12] One adult per household was selected to be interviewed using the Kish sampling table. From the 1,531 eligible respondents, 1,003 completed interviews were obtained (65.5 per cent) consisting of 63.4 per cent white/Anglos, 12.3 per cent blacks and 16.7 per cent Hispanics. Further details of the sample demographic characteristics are reported by Frerichs, et al.[13] ‡

The measures of restricted activity and utilization examined are similar to those used in the Health Interview Survey. The number of restricted activity days represents the total number of bed days and reduced activity days reported during the past two months. Thus, the number of restricted activity days ranged from 0 to 60. Information also was obtained about acute and chronic illness. For acute illness, the questionnaire item asked: "Thinking back to the past two months, were you sick or did you have any accidents or injuries at any time during the past two months? Please do not include any recurrent or continuing health problems that you might have all of the time." For chronic illness, the questionnaire asked: "Do you have a chronic, that is, a recurring or continuing health problem or condition; a health problem that you have all of the time or that makes you ill from time to time?" Up to three problems were recorded for both acute and chronic conditions. For each problem reported, respondents were asked to describe the medical problem and the number of times they contacted a doctor or other health professional about the problem during the past two months. Each reported symptom or problem was coded using the National Ambulatory Medical Care Survey Symptom Classification.[14]

---

‡ In this analysis, respondents were classified as Hispanic if they identified themselves a being Chicano or Mexican descent. Frerichs, et al.[13] also included in this category those who did not classify themselves as being of Chicano or Mexican descent, but who did classify themselves as being of "Other Spanish descent" (e.g., Cuban, Puerto Rican, etc.).

Measures of Access

In this analysis we have compared two of the leading "objective" indicators of access: the use–disability ratio as originally introduced by Aday,[4] and a measure of access similar to the one proposed by Chen.[3] The use–disability ratio is an aggregate measure calculated by dividing the number of physician contacts during a specified time period by the total number of disability days reported during the same period:

$$\text{Use–disability ratio} = \frac{\sum_{i=1}^{n} \text{MD}}{\sum_{i=1}^{n} \text{DD}}$$

where: n = number of persons in group reporting one or more disability days;

MD = number of doctor contacts;

DD = number of disability days (reduced activity days or bed days).

Since use–disability ratios are calculated for groups rather than individuals, it is difficult to test them statistically.§ Thus, inferences about "significant" differences on the use–disability ratio are usually based on a subjective evaluation as to what constitutes a meaningful difference (e.g., a difference of 5+ doctor contacts per 100 disability days).

The measure of access proposed by Chen is defined as the "proportion of people reporting a sign or symptom who acutally visited a physician":

$$p = \frac{n}{N}$$

where: n = number of persons reporting a sign or symptom who visited a physician;

N = number of persons reporting a sign or symptom.

At the individual level, scores on this measure can either be 0 or 1, depending on whether the person had symptoms and visited a physician (score = 1), or had symptoms and did not visit a physician (score = 0). Since this index uses the individual rather than the group as the statistical unit of analysis, significance testing is easily accomplished (e.g., by using χ^2 or F-statistics).

As originally proposed by Chen, the denominator of the index (i.e., the number of people reporting a serious sign or symptom) would be calculated from a predetermined list of health problems requiring medical attention. These problems would be identified by a panel of physician judges and included in the questionnaire in the form of a checklist. All respondents then would be asked whether they experienced a sign or symptom on the list and whether they saw a doctor about that problem.

---

§ The closest one can come to a direct statistical test of group level ratios involves using approximation methods like those discussed by Cochran.[15]

Since the LAHS was not specifically designed to examine Chen's index, we are not able to replicate it exactly as prescribed. In particular, our symptom data were not coded in sufficient detail to permit medical evaluations of their severity or appropriateness for medical care. Nonetheless, we have examined a variant of this index, which is defined as the proportion of people with a particular health problem that contacted a doctor.[#]

Results

Racial/Ethnic Differences in Access

Criticisms of the use–disability ratio have focused on the use of disability days as a measure of need and the fact that the ratio does not take the cause of the disability into account. Both of these problems became evident in our analysis of racial/ethnic differences in access to heal care. For example, Table 1 reports racial/ethnic differences on our two indicators of access: the use–disability ratio and the proportion of people with health problems who saw a doctor. As shown, blacks scored higher than white/Anglo and Hispanic respondents on both measures, suggesting that blacks had relatively greater access to health care. Notice, however, that while Hispanics scored lower than white/Anglos on the use–disability ratio by a factor of 2.5 doctor visits, their scores were essentially identical on the other measure of access.

This difference for the Hispanic subgroup illustrates an often-discussed problem with the use–disability ratio, namely, its sensitivity to subgroup differences in long-term disability. For example, 12.8 per cent of the Hispanics reported 60 days of restricted activity during the past two months, compared with only 8.3 per cent for white/Anglos and 9.8 per cent for blacks. Similarly, 14.5 per cent of the white/Anglos and 12.2 per cent of the blacks reported 30 or more restricted activity days, compared with 23.1 per cent of the Hispanics. This difference in long-term disability, of course, can disproportionately increase the denominator of the ratio for Hispanics—thereby reducing their overall use–disability ratio.

One strategy to minimize the effects of statistical outliers involves transforming the variable disability days (e.g., log transformation) before calculating the use–disability ratio. In so doing, we found that blacks were ranked highest (ratio using log disability days = 1.47), with white/Anglos (ratio = 1.31) and Hispanics (ratio = 1.34) recording somewhat lower scores. It is of interest to note that this ranking corresponds perfectly to that obtained from the other measure of access.

Another strategy to minimize the effects of outliers involves trimming or truncating the number of disability days at various cut-points. For example, one could truncate the number of disability days reported at 30+ so that the maximum number any one person can report would be 30. As shown in Table 2, we found very little difference among racial/ethnic subgroups when the scores were trimmed at 22+ days. However, as the number of disability days is allowed to increase (i.e., as we move from left to right in Table 2), racial/ethnic differences increase, with Hispanics showing progressively less access relative to their disability than either white/Anglo or black respondents. This pattern in the ratios indicates that the magnitude of the racial/ethnic differences are

---

[#] We also examined the proportion of health problems reported by an individual that resulted in a doctor contact. However, since most people in this survey reported a maximum of one health problem during the two month recall period, this variable was virtually identical to the one reported above (i.e., the proportion of people with health problems who contacted a doctor).

TABLE 1. Comparison of Two Measures of Access by Race/Ethnicity

	Use–Disability Ratio (past two months)*		Proportion of People With Health Problems That Reported a Doctor Contact (past two months)†	
Sample	20.3	(336)	0.52	(573)
Race/Ethnicity				
White/Anglo	20.7	(228)	0.49	(395)
Black	22.9	(41)	0.69	(64)
Hispanic	18.3	(39)	0.47	(73)

* Use/disability ratio × 100 equals the number of doctor contacts per 100 restricted activity days. This ratio is calculated only for those respondents who had at least one restricted activity day (N = 336).

† Calculated only for those people who reported at least one chronic or acute health problem during the past two months (N = 573). The denominator of the proportion equals the number of people with one or more health problems.

not uniform—partly because the subgroups differed in their rates of long-term disability.

The source of this problem became even more apparent upon closer examination of the data. That is, there were three Hispanics in this study reporting a total of 156 restricted activity days (60 + 60 + 36) and no doctor contacts. Neither the white/Anglo nor black subgroups had a similar subset of outliers. When we deleted these three people from the analysis, the use–disability ratio for Hispanics increased to 24.8, ranking them first among the three racial/ethnic groups. This particular ranking, however, illustrates a second problem with the use–disability ratio; namely, the failure of this ratio to consider differences in the etiology of the disability.

As column A in Table 3 shows, three major symptom categories were most likely to trigger a multiple doctor visit during the two-month recall period: acute and chronic musculoskeletal problems and cardiovascular problems. Moreover, as shown in column B, although proportionately fewer Hispanics with reduced activity reported cardiovascular problems, they did report a relatively high rate of acute musculoskeletal problems, and, by far, the highest rate of chronic musculoskeletal problems. Interestingly, when multiple doctor contacts were deleted from the numerator of the ratio, the obtained racial/ethnic differences were consistent with those found on the other indicator of access. That is, blacks were ranked first (ratio = 14.1), followed by Hispanics (11.1)

TABLE 2. Comparison of Use/Disability Ratios After Truncating Individual Reports of Restricted Activity Days by Race/Ethnicity*

	Trimmed 22+ Days	Trimmed 30+ Days	Raw Score
Race/Ethnicity			
White/Anglo (N = 228)	32.6	27.7	20.7
Black (N = 41)	33.7	30.5	22.9
Hispanic (N = 39)	31.8	26.2	18.3

* Use/disability ratio × 100 equals the number of doctor contacts per 100 restricted activity days. This ratio is calculated only for those respondents who had at least one restricted activity day during the past two months (N = 336).

and white/Anglos (10.9) with virtually identical scores. Thus, it would seem that Hispanics scored higher on the use/disability ratio (after deleting the three outliers), in part, because they reported more multiple doctor contacts than either black or white/Anglo respondents.

In the above analyses, the use–disability ratio was found to be sensitive to subgroup differences in long-term disability that could inflate the denominator of the ratio, and to differences in the etiology of the disability that could produce differential reports of multiple doctor contacts. The other indicator of access also has limitations, mainly because it uses health problems as the measure of need without taking into account variations in the severity of the problem. Thus, to examine the robustness of the racial/ethnic differences reported in Table 1, we compared these subgroups within categories of disability. As shown in Table 4, all of the comparisons continued to show greater use of health services by blacks than either white/Anglo or Hispanic respondents. Indeed, among the subgroups reporting 7+ restricted days, all 21 blacks made contact with a doctor, compared to 67 per cent and 79 per cent for white/Anglos (N = 81) and Hispanics (N = 14).

Up to this point, two well-known objective indicators of access have been compared: the use–disability ratio and the proportion of people with symptoms who contacted a doctor. A third index, defined as the proportion of people reporting disability days and had contacted a doctor, was also examined in this analysis. Using this index as a guide, blacks are ranked again as having the highest access, with 75.6 per cent reporting a doctor contact, compared with only 56.4 per cent for Hispanics and 52.2 per cent for white/Anglos ($p < 0.05$).

In yet another analysis, multiple regression techniques were used to estimate the number of doctor contacts reported by each ethnic group after statistically control-

TABLE 3. Proportion of People Reporting Particular Symptom Codes That Resulted in a Multiple Doctor Contact, and the Proportion Reporting These Symptoms by Race/Ethnicity

Symptom Codes†	A Proportion of Sample That Reported a Multiple Doctor Contact for Symptom		B Percentage of Racial/Ethnic Group With Reduced Activity That Reported Symptom*		
	%	N	White/Anglo	Black	Hispanic
Flu (313)	6.9	72	20.2	14.6	15.4
Acute respiratory (300–399)	9.3	118	29.8	24.4	17.9
Acute musculoskeletal (400–499)	35.4	65	9.2	22.0	20.5
Chronic musculoskeletal (400–499)	25.8	132	21.9	19.5	30.8
Cardiovascular (200–299)	34.3	102	13.6	12.2	7.7

* Percentages are based on number of people in group that reported at least one restricted activity day, and thus were included in the use/disability ratio. The number of white/Anglos was 228; blacks, 41, and Hispanics, 39.

† Symptoms were coded using the National Ambulatory Medical Care Survey Symptom Classification. Symptom codes are shown in parentheses.

TABLE 4. Proportion of People With Health Problems Who Made Contact With a Doctor by Race/Ethnicity and Level of Restricted Activity

Race/Ethnicity	Proportion of People With Health Problems Reporting a Doctor Contact* 1+ Restricted Days		3+ Restricted Days		7+ Restricted Days	
White/Anglo	0.52	(228)	0.61	(137)	0.67	(81)
Black	0.76	(41)	0.93	(29)	1.00	(21)
Hispanic	0.56	(39)	0.66	(29)	0.79	(14)
Significance level	$p < 0.05$		$p < 0.01$		$p < 0.01$	

* Number in parentheses represents the denominator of the percentage.

ling for differences in reports of illness and disability. As shown in Table 5, when no adjustments are made for racial/ethnic differences in reports of symptoms and disability, blacks averaged 1.47 doctor contacts, followed by white/Anglos (1.13) and Hispanics (0.76). When racial/ethnic differences in reports of symptoms and disability were controlled statistically, blacks were ranked first again—but by a somewhat larger margin ($p < 0.10$). As an aside, we might note that several cross-product interaction terms were also examined in this analysis, including race/ethnicity (dummy coded) by disability days. However, none of these interaction terms was statistically significant.

Although the number of cases was small, an examination of individual symptom codes showed that blacks were more likely to visit a doctor for respiratory and musculoskeletal symptoms (Table 6). A closer examination of these data, however, failed to uncover a common denominator that might explain this difference. For example, in one analysis we grouped all respondents reporting either respiratory or musculoskeletal problems (N = 278). Dummy variables representing blacks and Hispanics then were included in a multiple regression containing several other predictors, including respondent age, sex, total family income, educational and occupational ranking, number of restricted activity days in the past two months, regular source of health care (1 = no regular person or place, 2 = regular place but no regular person, 3 = regular person), and health in-

TABLE 5. Estimated Number of Doctor Contacts During the Two-Month Recall Period by Race/Ethnicity (N = 927)*

	White/Anglo	Black	Hispanic
Unadjusted scores	1.13	1.47	0.76
Adjusted scores	1.06	1.68†	1.36

* Number of doctor contacts was estimated using multiple regression. Race/ethnicity was dummy coded, with white/Anglo serving as the comparison group. The unadjusted scores represent the predicted scores for the three racial/ethnic groups from a regression containing only the two racial/ethnic variables (black vs. white/Anglo and Hispanic vs. white/Anglo) and the constant (white/Anglos). The adjusted scores represent the predicted scores from a regression containing the two racial/ethnic variables plus the number of restricted activity days (log of raw score + 1) and illnesses reported during the same recall period (i.e., the past two months). To estimate values for these latter two terms in the equation, their respective mean scores were used.

† $p < .10$.

TABLE 6. Proportion of People Who Contacted a Doctor About a Particular Symptom by Race/Ethnicity

System Category*	Proportion of People With Problem Who Made a Doctor Contact (past two months)		
	White/Anglo	Black	Hispanic
Respiratory (300–399)	33.7 (89)	54.5 (11)	30.0 (10)
Acute musculoskeletal (400–499)	52.4 (42)	83.3 (12)	70.0 (10)
Chronic musculoskeletal (400–499)	35.6 (87)	50.0 (14)	25.0 (20)
Cardiovascular (200–299)	56.0 (75)	71.4 (14)	75.0 (8)

Number in parentheses represents the denominator of the percentage.

* Symptoms were coded using the National Ambulatory Medical Care Survey Symptom Classification. Symptom codes are shown in parentheses.

surance (0 = does not have insurance 1 = has insurance). Results from this analysis showed that, controlling for all other variables in the equation, blacks were still significantly more likely to report a doctor visit than white/Anglo respondents ($\beta = 0.14$, $p < 0.05$). There were no significant differences between Hispanic and white/Anglo respondents ($\beta = 0.05$, NS).

Summary and Discussion

Two potential problems with the use–disability ratio were illustrated in these analyses. The first problem is that the ratio is sensitive to subgroup differences in chronic disability. Thus, when using this ratio for comparative purposes, it will be important to determine whether the subgroups differ in their rates of long-term disability. If the subgroups differ in this regard, then some type of adjustment should be made (e.g., "trimming" the distribution of disability days, deleting outliers, using a log transformation, etc.).

The second problem is that the ratio does not consider subgroup differences in the cause of the disability. Certain sources of disability (e.g., acute vs. chronic illness) may be more or less likely to result in a doctor contact. Moreover, once contact with a doctor has been made, certain sources of disability may be more or less likely to result in multiple doctor contacts (e.g., cardiovascular vs. respiratory problems). If the sample size is large enough, it therefore may be advisable to examine use–disability ratios separately for acute vs. chronic illness, or perhaps within categories of symptoms. Of course, making such comparisons presumes that one can link reports of restricted activity days with reports of doctor visits for particular symptoms.

Given these problems with the use–disability ratio, one might question the value of using this index at all. We found the use–disability ratio to be somewhat awkward to use, since a considerable amount of preliminary work had to be done to determine whether nonaccess factors were influencing the results. However, if subgroups are comparable in their illness profiles (e.g., acute vs. chronic illness, reports of particular symptom codes, etc.) and rates of long-term vs. short-term disability, or are made comparable by "trimming," deleting outliers, etc., then the use–disability ratio would seem to represent a reasonable measure of access. That is, all things being equal, subgroups reporting higher rates of use relative to their disability may indeed have greater access than subgroups scoring lower on the use–disability ratio. Yet, perhaps the most serious limitation of the use–disability ratio concerns the unit of analysis. Because the use–disability ratio is a group level variable, subgroup differences are not easy to test. Indeed, of all the analyses we conducted, the results obtained from the multiple regression were the most informative—offering significance testing with multivariate controls, and estimates of change in the dependent variable (i.e., number of doctor visits) attributable to other measured variables of interest (e.g., disability days, income, race/ethnicity, etc.).

Finally, the inability to explain why blacks were more likely to report a doctor contact for respiratory and musculoskeletal problems should be mentioned. This difference occurred independently of standard sociodemographic variables (e.g., age, sex, income, etc.), as well as the more policy-relevant variables pertaining to regular source of care and health insurance. Perhaps the answer lies in racial/ethnic differences in the dimensions of access identified by Penchansky and Thomas[2] (i.e., availability, accessibility, accommodation, affordability and acceptability), or in the type of health-related advice offered by friends, family, spouse, etc.[16–19] This ambiguity illustrates the need to link objective indicators of access with factors hypothesized to influence use of services. Thus, in the final analysis, we again are reminded that informed policy decisions require an assessment of use relative to need, as well as an understanding of the variables responsible for the differential rates of use. Given these two requirements, it would seem that the most useful indicators of access will be those that make the individual, rather than the group, the main focus of the analysis.

References

1. Aday LA, Andersen R. Access to medical care. Ann Arbor, MI: Health Administration Press, 1975.

2. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. Med Care 1981;19:127.

3. Chen MK. Measuring need for health services: a proposed model. Med Care 1979;17:210.

4. Aday LA. Economic and noneconomic barriers to the use of needed medical services. Med Care 1975;13:447.

5. Aday LA. Letter to the editor. Med Care 1976;14:717.

6. Andersen R. Health status indices and access to medical care. Am J Public Health 1978;68:458.

7. Aday LA, Andersen R, Fleming GV. Health care in the United States: equitable for whom? Beverly Hills: Sage Publications, 1980.

8. Montiero LA. Monitoring health status and medical care. Cambridge: Ballinger Publishing Company, 1976.

9. Marcus AC, Reeder LG, Jordan LA, Seeman TE. Monitoring health status, access to health care, and compliance behavior in a large urban community: a report from the Los Angeles Health Survey. Med Care 1980;18:253.

10. Nathanson C. Social roles and health status among women: the significance of employment. Soc Sci Med 1980;14A:463.

11. Fox PD, Bice TW. Socioeconomic status and use of physician services revisited. Med Care 1976;14:714.

12. Sumner J. The 1976 LAMAS frame and master sample: technical description. Los Angeles: Institute for Social Science Research, 1976.

13. Frerichs RR, Aneshensel CS, Clark VA. The prevalence of depression in Los Angeles County. Am J Epidemiol 1981;113:691.

14. National Center for Health Statistics. The national ambulatory medical care survey: symptom classification. USDHEW, Public Health Service, Vital and Health Statistics, U.S. Department of Health, Education and Welfare, Series 2:63, 1974. (DHEW Publication No. 74-1337.)

15. Cochran WG. Sampling techniques. New York: John Wiley and Sons, 1977.

16. McKinlay JB. Some approaches and problems in the study of the use of services: an overview. J Health and Soc Behav 1972;13:115.

17. Langlie JK. Social networks, health beliefs and preventive health behavior. J Health Soc Behav 1977;18:244.

18. Safer MA, Tharps, QT, Jackson TC, Leventhal H. Determinants of three stages of delay in seeking care at a medical clinic. Med Care 1979;17:11.

19. Berkanovic E, Telesky C, Reeder S. Structural and social psychological factors in the decision to seek medical care for symptoms. Med Care 1981;19:7.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Does Income Affect Mortality? An Analysis of the Effects of Different Types of Income on Age/Sex/Race-Specific Mortality Rates in the United States

JACK HADLEY, PH.D.,* AND ANTHONY OSEI, M.A.†

This article explores the question of whether or not higher incomes are associated with lower mortality rates. Some recent research on this issue has suggested that income either has no effect on or may even be positively correlated with mortality rates. By contrast, earlier studies consistently found a negative relationship—higher income (or economic status) was generally associated with lower mortality rates. This paper extends the prior research in two significant ways. First, the issue is analyzed separately for eight adult and four infant age/sex/race-specific population cohorts. Second, total family income is broken down into several components to investigate whether different types of income have differential effects on mortality rates. In addition, the problem of untangling the joint effects of education and income on mortality also is explored. The results tend to support the hypothesis that higher income is associated with lower mortality rates. However, the magnitude of the impact of income is small, although it is consistently larger for infants than for adults.

ALL COUNTRIES VALUE good health as an important part of individual welfare. It is also a significant factor in the level of economic activity, influencing both labor force participation and worker productivity. One variable that is used frequently to measure health is the mortality rate. It generally is agreed that the recent decreases in mortality rates in the United States have been due to several factors: scientific and technologic advances that increased the supplies of commodities contributing to good health; greater educational attainment, making people better health "producers"; increased use of improved quality medical care; and higher real incomes. For example, more than 15 years ago the American Medical Association stated, "Medical science does not seek major credit for the improvements in the health level during the past 25 years. Certainly, our standards of living and higher education levels have contributed substantially to the betterment of the health level in the United States."[1]

Such statements partially explain why health services researchers recently have tended to focus on socioeconomic conditions to explain mortality differentials. This article continues to work in this vein

---

* Senior Research Associate, The Urban Institute, Washington, D.C.

† Research Assistant, The Urban Institute, Washington, D.C.

Supported by Grant No. 5-R01-HS02790 from the National Center for Health Services Research, OASH, USDHHS.

The findings and opinions expressed in this paper are solely those of the authors and do not necessarily reflect the positions of The Urban Institute or the Department of Health and Human Services.

0025-7079/82/0900/0901/$01.50 © J. B. Lippincott Co.

by focusing on the relationship between income and mortality rates. Specifically, is income a significant factor in explaining geographic variations in age- sex- race-specific mortality rates? This question is particularly interesting in this time of "budget crisis." With all levels of government concerned about the size and growth of transfer payments, which constitute a substantial portion of the incomes of certain population subgroups, it is important to ask what consequences cuts in transfer payments might have on the health of the population. Will mortality rates increase, decrease or level off with changes in income?

The existing evidence on this issue is conflicting. Some studies imply that death rates tend to be correlated negatively with income, while others postulate an insignificant or even positive correlation. Although the factors underlying the latter relationship are not well understood, several hypotheses have been offered to explain this apparent paradox. They include the existence of a ceiling on the ability of standards of living to improve health which already has been reached by certain population groups; positive income-elasticities of demand for certain goods whose consumption adversely affects health; and the existence of certain nonconsumption factors, associated with higher income occupations, which have deleterious health consequences (psychological pressure and tensions in managerial occupations, for example).[2,3]

The primary objective of this paper is to extend the research on the general question "Is income a significant factor in explaining geographic variations in mortality rates?" This is done in two ways. First, income is separated into various components to investigate whether different types of income have different effects on mortality rates. Second, the analysis is based on detailed mortality and sociodemographic data, which are aggregated into eight adult and four infant age- sex- race-specific cohorts. Thus, these data are more detailed than those used by many earlier studies.

Prior Research

Much of the earlier work done on variations in mortality rates generally can be classified into two groups: a) epidemiologically oriented studies, basically concerned with identifying differences in mortality rates for population groups classified by different socioeconomic criteria and b) those studies concerned with explaining observed differences in mortality rates by applying more formal theoretical models and multivariate analytic techniques. Among the former are works by Altenderfer, Antonovsky, and Lerner and Stutz.[4,5,6] (See Kitagawa and Hauser for a thorough review of this extensive literature.[7]) These researchers investigated various geographic units, time periods, and mortality rates. All found an inverse relationship between socioeconomic status and mortality. This basic finding is summarized by Antonovsky's observation that "Despite the multiplicity of methods in the 30-odd studies cited, and despite variegated populations surveyed, the inescapable conclusion is that class influences one's chances of staying alive. Almost without exception, the evidence shows that classes differ on mortality rates."[5] Two things to be noted about these studies, however, are the aggregate nature of the mortality indices used and the crude and unrefined measures of income employed.

Turning to the second group of studies, Auster et al., following earlier work by Fuchs, proposed a health production model to examine the relationship between age-adjusted death rates, measures of medical care consumption and a number of environmental variables, including income.‡[2,8] They found that income is related positively to death rates and that this relation remains the same after dropping education from their analysis. They speculated that above some point, rising income may be associated with occupational conditions (such as stress, lack of exercise) and behavior changes (such as in diets, increased auto travel) that have adverse effects on mortality.

---

‡ The authors in this study were more concerned with estimating the impact of medical care on mortality rates than estimating the effects of income.

Although this study made an important contribution in this area, it has at least two shortcomings. First, as the authors themselves point out, it would have been preferable to estimate separate age- sex- race-specific production functions, since earlier work suggests that there are important differences in the factors affecting specific mortality rates. A second problem is the use of the state as the unit of analysis, since extensive intrastate heterogeneity may be masked by the use of state averages.

Silver extended the work of Auster et al., by disaggregating the mortality data by race and sex and estimating the relationships for both states and SMSAs.[3] In general, the results did not tend to support the Auster et al. findings of a positive relationship between household income and mortality (although the state-level estimates for white males were positive and significant). However, as the author cautions, these findings are derived from regressions which exclude measures of schooling, physicians per capita and fertility on the grounds of multicollinearity. When income was divided into its labor and nonlabor components, the effect of labor income tended to be positive (or less negative) while the coefficient of nonlabor income was consistently negative. The positive effect of labor income was strongest for white males, which suggests that for this subset of the population, the factors associated with increasing income may have deleterious side effects.

Grossman, in another study, concluded "Perhaps the most striking finding in this case is that health has a negative income elasticity. One explanation of this result is that the income elasticities of the detrimental inputs in the health production function exceed those of the beneficial inputs, . . ."[9] Orcutt et al. estimated variations in death rates across county groups.[10] Their analysis removed the effects of person-specific factors (age, sex, race, etc.) and attributed any residual variations to other factors such as income. They concluded that higher incomes are correlated with lower death rates. Finally, Margaret Reid examined the conflicting evidence on the relationship between death rates and income.[11] Her analysis, which covered white populations of states for 1970, found little or no evidence to support the view that death rates tend to be positively correlated with income. Although the author offered no concrete explanation as to why her income and education coefficients were not as strongly negative as those of earlier studies, she did suggest that use of age- sex-race-specific death rates as opposed to age-adjusted death rates might lead to a better understanding of the income-mortality paradox.

Empirical Methods

This section describes the empirical model that underlies the analysis. Specific issues discussed are the choice of a unit of analysis, the choice of mortality as the dependent variable, data sources, econometric issues (simultaneous equation estimation), specification of the estimating equation and variable definitions.

A. Unit of Analysis and Data

The geographic unit of analysis for this study is the county group, as defined by the Bureau of the Census. The entire country is divided into areas and subareas, each of which has a population of at least 250,000 people.[2] § (County groups may cross state boundaries.) There were 407 identified areas and subareas in 1970.

---

§ Each area is an extension of the SMSA concept—an urban center and surrounding counties tied to that center by trade and commuting patterns. In rural areas which do not contain an SMSA, the urban center is defined as a city with 20,000 to 50,000 population. Identified subareas within each county group are SMSAs or components of SMSAs with populations greater than 250,000 people.

Use of the county group as the unit of analysis offers more observations than use of either states or SMSAs. In addition, there should be less intra-area variation in the data in a county group than in a state. At the same time, county groups provide information on rural areas, since the presence of a central city of more than 50,000 people is not required to construct a county group. Compared to individual counties, the county group contains a much larger population for analysis and is not as likely to be affected by the border-crossing problem, particularly since the county group can cross state boundaries.

Within each county group, the population was divided into 12 cohorts consisting of three age groupings, 0 to 1, 45 to 64, and 65 and older, for both sexes by race (white and black). (These cohorts account for about 90 per cent of all deaths.) The main reason for disaggregating the population in this way is to control for the effects of variations in biological and sociocultural characteristics that may be associated with age, sex and race but are difficult to measure directly. The most obvious factor is natural changes in health stock usually associated with aging. Similarly, attitudes about health and medical care, the effects of racial discrimination, and inherited physical traits may be other factors that are difficult to observe but can be controlled partially by investigating age- sex- race-specific population cohorts.

Sociodemographic characteristics of each cohort were constructed by aggregating individual data from a 2 per cent sample of the 1970 census to the county group level. In order to increase the reliability of the data, a minimum of 2,000 people in a particular cohort was required for inclusion in the analysis. (For the infant analysis, a minimum of 20 deaths in each sex-race category was required.) Finally, the dependent variable is the mortality rate from all causes of death, disaggregated by age, sex and race. These rates are based on deaths over the period 1968 through the first half of 1972 in order to lessen random fluctuations in death rates.[12]

B. Empirical Specification

Preliminary analyses, reported in an earlier article, suggested the use of the log-linear functional form.[13] Thus, the dependent variable in the specifications reported below is the log of mortality from all causes. The independent variables attempt to capture the effects of initial health status, environmental conditions and medical care consumption. Table 1 lists these variables and reports their mean values for each adult cohort. Table 2 reports the mean values of the infant cohort variables, which describe characteristics of births and of women of child-bearing age (14 to 44).

The basic equations were estimated with several alternative definitions of income: Total Family Income (TFI)—the sum of the dollar amounts of wages, salary income, net nonfarm and farm self employment and other income of all household members 14 years old and over; Earned Income (EI)—the sum of wages, salaries, commissions, bonuses or tips from all jobs and self-employment, professional and farm income before deductions (excluding military bonuses, reimbursements and pay in kind); Unearned Income (UI)—total family income less earned income (TFI-EI); Social Security Income (SSI)—U.S. government payments to retired persons, to dependents of deceased insured workers or to disabled workers (excluding Medicare reimbursements); Welfare Income (WI)—amounts received from federal, state and local public programs such as AFDC, old age assistance, general assistance, and aid to the blind or totally disabled; Transfer Income (TI)—the sum of SSI and WI; and Other Unearned Income (OUI)—interest, dividends, veterans' payments, retirement pensions from private employers, unions and other governmental agencies, rental

TABLE 1. Mean Values of Regression Variables, Adult Cohorts

	WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+
LALLR—Log deaths from all causes	2.63	4.33	3.13	4.28	1.88	3.83	2.63	3.94
AHG—Average highest grade	12.61	10.48	9.72	7.21	12.72	10.98	10.46	8.11
MTJA—Mean temperature jan	34.94	34.94	40.08	43.03	39.94	34.94	40.32	41.83
PWS—% Public water system	36.25	36.43	40.00	36.95	37.43	39.16	40.80	40.17
EXALCIR*—Excess alcohol and cirrhosis deaths	0.0001	−0.0057	0.0026	−0.248	0.0003	−0.0057	−0.0043	0.0054
AVA—Average age	54.28	74.64	54.21	73.52	54.33	74.80	54.20	74.07
USB—Per cent U.S. born	95.59	86.93	98.82	98.51	95.33	89.18	98.94	98.94
MIGR—Per cent migrants	2.97	2.18	1.79	1.23	2.89	2.31	1.84	1.23
EXFLU*—Excess flu deaths	0.0020	0.0017	0.0080	0.0003	0.0006	−0.0002	0.0072	0.0014
NILF—Per cent not in labor force	N.A.	N.A.	N.A.	N.A.	36.17	73.76	27.03	66.12
KIDS—Average no. children born	N.A.	N.A.	N.A.	N.A.	2.51	2.81	3.07	3.13
DISABLH†—Per cent work-limiting disability	10.13	N.A.	12.60	N.A.	8.41	N.A.	13.24	N.A.
LMCDREXP‡—Log medicare expenditures per enrollee	5.68	5.68	5.69	5.65	5.68	5.68	5.68	5.64
MAR—Per cent married	88.09	72.41	73.59	59.81	75.72	35.27	56.55	27.22
CIGS§—Cigarettes sold per person (00s)	23.78	23.78	23.96	23.72	23.78	23.78	23.85	23.58
VET—Per cent veteran	25.78	13.16	20.55	10.28	N.A.	N.A.	N.A.	N.A.
WORK68—Per cent last worked after 1968	10.89	17.20	13.25	15.75	N.A.	N.A.	N.A.	N.A.

* Defined as the difference between actual and predicted deaths.

† Predicated value from a first-stage regression.

‡ Values for white cohorts are predicted from the first-stage regression of Medicare expenditures per enrollee with all of the independent variables in the model.

§ Defined for each cohort's state.

N.A. = Not applicable/available.

TABLE 2. Mean Values of Regression Variables, Infant Cohorts*

	WF 14–44	BF 14–44
LMIR—Log male infant death rate	2.97	3.55
LFIR—Log female infant death rate	2.70	3.35
AHG—average highest grade	13.52	12.20
MAR—Per cent married	87.35	57.14
NILF—Per cent not in labor force	27.18	19.44
USB—Per cent U.S. born	95.84	98.78
KIDS—Average no. of children born	2.83	3.29
SPAN—Per cent Spanish	4.09	1.32
CIGS†—Cigarettes sold per person (00s)	23.76	23.63
LMCDREXP—Log medicare expenditures per enrollee	5.69	5.68
LEMIR--Log expected male infant death	3.23	3.91
LEFIR--Log expected female infant death	2.96	3.68
PMBHFM‡—Per cent male births high risk mothers	6.70	7.90
PFBHRM‡—Per cent female birth high risk mothers	6.80	8.00
PMBIH—Per cent male births in hospitals	99.70	95.70
PFBIH—Per cent female births in hospitals	99.70	95.50

* AHG, MAR, NILF, USB, ANKB, and SPAN are based on data from the 1970 Census for women of child-bearing age (14 to 44). Data on births and deaths are from vital statistics records.

† Defined for the state.

‡ Mothers younger than 15 and older than 34.

income, workman's compensation, alimony, private welfare payments and contributions from persons who are not members of households.

Table 3 reports the means of these variables for each cohort. As is well known, whites have higher family incomes than blacks in each age-sex group, by factors ranging from 60 to 85 per cent. White-black differences are much smaller for transfer incomes from public sources (Social Security payments and welfare income). In fact, blacks have higher levels of transfer income than whites except in the two elderly age groups. Differences are much greater, on the other hand, for transfer and unearned income from private sources (OUI), with whites having from two to six times more income from these sources than blacks.

All equations were estimated by ordinary least squares (OLS). In addition, the adult-cohort equations also were estimated by a simultaneous equation method, instrumental variable (IV) estimation. The latter approach is justified on the grounds that for adults, both medical care and income may be simultaneously determined with health. Thus, each of the income variables and the medical care use variable were replaced by predicted values generated by regressing income and medical care against all of the exogenous variables specified by a larger model of health production.[13]

Preliminary results indicated that the OLS and IV estimates of the coefficients of the income variables were very similar. This suggests that, at least for these data, simultaneous determination of mortality rates and the various income measures may not be a problem. The IV estimates of the coefficients of the medical care variable, however, did differ from the OLS estimates for the white cohorts, but not for the black cohorts. Based on these preliminary results, therefore, the income coefficients reported below for adults are based on OLS estimation of the black cohorts' mortality regressions and IV estimation of the white cohorts' equations (with medical care use replaced by its predicted value).

As noted previously, one of the explanations offered for the income–mortality

TABLE 3. Mean Values of Income Variables, by Cohort

	WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+	WF, 14–44	BF, 14–44
TFI—Total family income	$11,791	$6,570	$7,272	$3,941	$10,376	$5,863	$6,226	$3,811	$11,691	$6,821
EI—Earned income	8,242	1,638	4,377	857	2,060	405	1,559	303	1,971	2,012
UI—Unearned income	3,463	3,808	2,733	2,225	8,229	4,732	4,466	2,842	9,720	4,609
SSI—Social security income	83	1,089	110	730	83	666	101	468	18	35
WI—Welfare income	3	35	52	129	4	59	99	198	16	130
TI—Transfer income	86	1,125	162	859	87	725	200	666	34	165
OUI—Other unearned income	603	1,327	143	268	265	673	627	113	104	39

Definitions: *Total family income*—The sum of the dollar amounts of wages, salary income, net non-farm and farm self employment, and other income of *all* household members 14 years old and over. *Earned income*—Includes wages, salaries, commissions, bonuses or tips from all jobs and self employment, professional, and farm income before deductions (excluding military bonuses, reimbursement, and pay in kind). *Unearned income*—Total family income less earned income. *Social Security income*—Includes U.S. government payments to retired persons, to dependents of deceased insured workers, or to disabled workers (excluding Medicare Reimbursements). *Welfare income*—Includes amounts received from federal, state and local public programs such as AFDC, old age assistance, general assistance, and aid to the blind or totally disabled. *Transfer income*—The sum of S.S.I. and W.I. *Other unearned income*—Includes interest, dividends, veterans payments, retirement pensions from private employers, unions and other governmental agencies, rental income, workman's compensation, alimony, private welfare payments, and contributions from persons not members of households.

paradox is that the underlying relationship is nonlinear, i.e., mortality rates fall as income rises, but in successively smaller decrements. Eventually, on the "flat" portion of the curve, further increases in income will have no impact on mortality rates. This hypothesis can be tested by adding an income-squared variable to the model. Its expected sign is positive. Preliminary tests of this nonlinearity hypothesis suggested that the functional form reported below is superior.[‖] This result also suggests that differences in the variance of the income distributions of the cohorts within a county group can be ignored. If the income-mortality relationship is approximately linear over the relevant range, then the experiences of persons in the upper and lower tails of the income distribution will offset each other.[¶]

Finally, several earlier studies have emphasized the difficulty of untangling the effects of income on mortality rates from those of education.[2,3,9] This problem is addressed by estimating each equation both with and without the education variable in order to gauge the strength of the interrelationship between these factors. In all, 14 specifications are estimated. The last seven are essentially the same as the first seven, except they exclude education. The seven main specifications differ with respect to the definition of income.

Results

The coefficients of the adult cohorts' income variables (from the equations with education included) are reported in Table 4. Although the underlying equations explain exceedingly high proportions of the variance in mortality from all causes (R^2s vary around 0.9), these coefficients do not provide an unambiguous resolution of the income–mortality issue. The frequency of negative coefficients appears to tilt the balance in favor of an inverse relationship, but a number of the coefficients are not statistically significant. In Specification 1, for example, the coefficients of TFI, total family income, are negative for five of the cohorts, but statistically significant only for the two middle-aged male cohorts.

When total family income is decomposed into its earned (EI) and unearned (UI) components, a different pattern emerges. Earned income coefficients are negative and significant for three of four middle-aged cohorts and positive for three of four elderly cohorts. For unearned income however, the signs of the coefficients are reversed between middle-aged and elderly cohorts. In general, the pattern is that EI has a negative effect for the younger cohorts and positive effect for elderly cohorts. This should not be too surprising, however, since this variable partially captures the influence of initial health, which affects both labor force participation and mortality rates. People in good health and of greater economic means generally have higher earned incomes while of working age, and higher unearned income (and thus less need to work) while of retirement age.

Specifications 3 and 4 estimate the impact of earned and unearned income separately. The results suggest no significant qualitative or quantitative differences from Specification 2 on the impact of both EI and UI on mortality rates either through joint or independent estimation. The coefficients of both earned and unearned in-

---

[‖] In seven of the eight adult cohorts, both the linear and squared terms entered with implausible signs, positive and negative; respectively. This pattern implies that mortality rates initially increase as income increases, but eventually decline in successively larger decrements. A double-log specification also was tested, but its implications were essentially the same as those reported below.

[¶] Another factor which lessens the need to account explicitly for the variance of income is stratification of the population into age-, sex- race-specific cohorts. As Table 3 indicates, intercohort variation is larger than intracohort variation. The mean coefficient of variation across cohorts is 0.31. When the cohorts are pooled, the coefficient of variation jumps to 0.45.

come are essentially identical to those estimated by Specification 2.

The remaining specifications focus on different components of unearned income. As noted above, UI includes the earnings of other family members as well as rental and capital income, and various transfer payments from public and private sources. Looking first at public transfer programs, Specification 5 examines the impact of transfer income. In general, the results are ambiguous. Although five of the coefficients are negative, only two are statistically significant. However, two of the three positive coefficients are also statistically significant. These results do not lend themselves to meaningful interpretation.

Decomposing public transfer payments into Social Security payments and welfare payments helps somewhat to clarify the ambiguity of the TI coefficients. In particular, seven of the eight coefficients of welfare income are negative, with four being statistically significant. Conversely, six of the SSI coefficients have positive signs, with three being statistically significant. Since SSI includes payments to deceased insurees' spouses, who generally have higher mortality rates than either married or single people,[14] and to the disabled, it may be that this variable is strongly associated with poor health status. The inability of the instrumental variable estimates of Specifications 5 and 6 to detect this relationship may have been the result of inadequate instrumental variables for TI, SSI and WI, rather than the absence of any relationship between health and these variables.

Finally, the last specification focuses on unearned income from capital sources and from private pension plans. The qualitative results for this variable are very similar to those for pure public transfer payments, WI. The coefficients are negative for seven of the eight cohorts and five of these also are statistically significant. The strength of this finding may be due to the fact that this

TABLE 4. O.L.S. Income Coefficients, Adult Cohorts (Education Included)

Income Specification		WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+
1.	TFI	−0.0006*	−0.0003	−0.0016†	0.0003	−0.0003	NS	−0.0003	−0.0003
2.	UI	0.0006	−0.0010‡	0.0006	−0.0009	0.0001	0.0005	0.0001	−0.0001
	EI	−0.0010†	0.0007	−0.0033†	0.0032*	−0.0032*	−0.0095†	0.0013	0.0007
3.	UI	0.0003	−0.0007‡	−0.0002	−0.0005	−0.0002	0.0001	0.0002	−0.0001
4.	EI	−0.0010†	0.0002	−0.0032†	0.0029*	−0.0030*	−0.0084†	0.0014	0.0007
5.	TI	−0.0106‡	−0.0021	0.0060	−0.0039	0.0117‡	0.0195†	−0.0031	−0.0118*
6.	SSI	−0.0073	0.0009	0.0143‡	0.0001	0.0110‡	0.0283†	0.0009	−0.0009
	WI	−0.0272‡	−0.0247†	−0.0087	−0.0223†	0.0158	−0.0118	−0.0058	−0.0023†
7.	OUI	−0.0028‡	0.0037†	−0.0040	0.0041‡	−0.0148†	−0.0052†	−0.0216*	−0.0200†

NS, No significant digits.
* Significant at the 95 per cent confidence level.
† Significant at the 99 per cent confidence level.
‡ Significant at the 90 per cent confidence level.

form of unearned income is least likely to be related to health status. As such, it may be the income component that is most closely correlated with wealth and permanent income.

The results for the infant cohorts (from the specification including education) are presented in Table 5. In general, the evidence suggests a fairly consistent negative relationship between mortality and income for all seven specifications. Coefficients are similar in magnitude for female and white male infants, but generally smaller in magnitude for black male infants. Both mothers' earned income and unearned income from all sources, which consists primarily of spouses' and other family members' earnings, have uniformly negative signs and are statistically significant in all but one case. The impact of public transfer income, which is a very small component of total family income on the average, is ambiguous. Finally, private transfer income, which makes up less than 1 per cent of total income, has negative coefficients for three of the four cohorts.

Education and income are highly correlated, as is well known. Thus, including an education variable in the regression equations increases the unreliability of the estimated income coefficients and makes their interpretation more ambiguous. Analysis of the simple correlation coefficients between TFI and education provides evidence to support this hypothesis. The correlation coefficients vary from 0.184 (elderly white males) to 0.766 (middle-aged white males) for the adult cohorts and 0.561 (blacks) to 0.770 (whites) for the infant cohorts (women of childbearing age). However, the apparent presence of a relatively high degree of multicollinearity is not sufficient reason to follow Silver and exclude education from our equations.[3] First, there is ample evidence in the literature to suggest the relevance of education as an important factor in explaining variations in mortality rates. Second, the exclusion of a relevant variable such as education may lead to serious error: not only may the estimated coefficients be biased, but inferences based on them also will be inaccurate. As Johnston suggests, "Data and degrees of freedom permitting, one should err on the side of including variables in the regression analysis rather than excluding them."[15]

In order to gauge the impact of multicollinearity on the results, we also report the income coefficients from equations that exclude the education variable. These results are presented in Tables 6 and 7. In

TABLE 5. O.L.S. Income Coefficients, Infant Cohorts (Education Included)

Income Specification		Males		Females	
		Blacks	Whites	Blacks	Whites
1.	TFI	−0.0007	−0.0016*	−0.0023†	−0.0019*
2.	UI	0.0012	−0.0011*	−0.0014	−0.0017*
	EI	−0.0063‡	−0.0085*	−0.0048	−0.0043‡
3.	UI	−0.0001	−0.0014*	−0.0024‡	−0.0019*
4.	EI	−0.0051‡	−0.0106*	−0.0063‡	−0.0076*
5.	TI	0.0002	0.0110	−0.0036	0.0015
6.	SSI	−0.0243	0.0197‡	−0.0406‡	0.0068
	WI	0.0114	0.0008	0.0128	−0.0047
7.	OUI	−0.0093	−0.0136‡	0.0142	−0.0112‡

* Significant at the 99 per cent confidence level.
† Significant at the 95 per cent confidence level.
‡ Significant at the 90 per cent confidence level.

general, the signs of the coefficients suggest an inverse relationship between income and mortality. As expected, most of the coefficients are biased negatively.# Only one coefficient (elderly black males) appears to have a positive bias. For total family income (TFI) the average percentage increase in absolute value for the negative coefficients is about 50 per cent.

A similar analysis of the effect of income on the education coefficients can be conducted by estimating regressions that include and exclude the income variables. The result of such an analysis for all cohorts is presented in Table 8, where the income variable in question is total family income. As can be gleaned from the table, education coefficients, with total family income excluded, are, in general, negatively biased and larger in absolute value than those with family income included. Nine of 12 coefficients increased in absolute value, on the average by about 60 per cent, with the maximum and minimum changes occurring for middle-aged white females and elderly black males, respectively. In general, education has a consistently negative association with mortality rates.

Conclusions

Several research and policy issues were raised at the outset of this paper. Is income a significant factor in explaining geographic variations in adult mortality rates? What are the differential effects (if any) of the various components of income? What is the effect of "pure" unearned income on mortality rates?

These questions were addressed through estimation of a health production function that used an inverse indicator, the mortality rate from all causes, as the dependent variable. The primary empirical

TABLE 6. O.L.S. Income Coefficients, Adult Cohorts (Education Excluded)

Income Specification		WM, 45–64	WM, 65+	BM, 45–64	BM, 65+	WF, 45–64	WF, 65+	BF, 45–64	BF, 65+
1.	TFI	−0.0011*	−0.0003	−0.0016*	−0.0004	−0.0002	−0.0004†	NS	−0.0005
2.	UI	0.0005	−0.0006‡	0.0009	−0.0009	0.0001	−0.0003	0.0002	−0.0005
	EI	−0.0017*	0.0007	−0.0034*	0.0007	−0.0026†	−0.0034†	−0.0004	−0.0014
3.	UI	−0.0002	−0.0005†	NS	−0.0008	−0.0001	−0.0004†	0.0001	−0.0005
4.	EI	−0.0017*	0.0002	−0.0032*	0.0004	−0.0023†	−0.0039†	−0.0003	−0.0016
5.	TI	0.0029	−0.0033†	0.0023	0.0032	0.0128‡	0.0126*	−0.0057	−0.0017
6.	SSI	0.0075	−0.0029	0.0117†	0.0059†	0.0121†	0.0203*	−0.0020	−0.0014
	WI	−0.0196	−0.0095†	−0.0138†	−0.0095	0.0162	0.0030	−0.0081	−0.0020
7.	OUI	−0.0051*	−0.0028*	−0.0059†	0.0048†	−0.0131*	−0.0044*	−0.0225‡	−0.0083

NS, No significant digits.
* Significant at the 99 per cent confidence level.
† Significant at the 90 per cent confidence level.
‡ Significant at the 95 per cent confidence level.

# On theoretical grounds, if education is hypothesized to be positively related to income but negatively related to mortality, then omitting education from the regression equation should negatively bias the income coefficients.

TABLE 7. O.L.S. Income Coefficients, Infant Cohorts (Education Excluded)

Income Specification		Males		Females	
		Blacks	Whites	Blacks	Whites
1.	TFI	−0.0011	−0.0021*	−0.0029†	−0.0026*
2.	UI	0.0012	−0.0016*	−0.0015	−0.0024*
	EI	−0.0067†	−0.0094*	−0.0060‡	−0.0053‡
3.	UI	−0.0007	−0.0020*	−0.0032†	−0.0026*
4.	EI	−0.0053†	−0.0133*	−0.0078*	−0.0111*
5.	TI	−0.0035	0.0146‡	−0.0099	0.0061
6.	SSI	−0.0264	0.0128	−0.0443‡	−0.0008
	WI	0.0066	0.0167	0.0050	0.0142
7.	OUI	−0.0114	−0.0290*	0.0091	−0.0315*

Income and education are based on data for women of child-bearing age.

* Significant at the 99 per cent confidence level.

† Significant at the 95 per cent confidence level.

‡ Significant at the 90 per cent confidence level.

extension offered by the work reported here was estimation of these functions with seven more refined measures of income entering the specifications. Separate relationships were estimated for 12 population cohorts consisting of three age groups, 0 to 1, 45 to 64, and 65 and older, for both sexes by race (white and black).

Based on the results reported above, the answer to the first question would seem to be a qualified yes. Analysis of all cohorts and specifications reveals a negative relationship in general. Although a number of the estimated coefficients were statistically insignificant, particularly for adults, examination of coefficients from variants

TABLE 8. Coefficients of Education, All Cohorts*

Cohorts	Pearson Correlation Coefficient†	Income Included	Income Excluded
Adults			
WM, 45–64	0.766	−0.0205‡	−0.0239‡
WM, 65+	0.533	−0.0034	−0.0122‡
BM, 45–64	0.702	−0.0132§	−0.0203‡
BM, 65+	0.533	−0.0104§	−0.0099
WF, 45–64	0.631	−0.0165‖	−0.0150‡
WF, 65+	0.184	−0.0203‡	−0.0233
BF, 45–64	0.600	−0.0382‡	−0.0418‡
BF, 65+	0.431	−0.0126	−0.0099
Male Infants			
Whites	0.770	−0.0522‡	−0.0754‡
Blacks	0.561	−0.0180	−0.0252
Female Infants			
Whites	0.770	−0.0651‡	−0.0945
Blacks	0.560	−0.0242	−0.0479

* Income = total family income.

† Income and education.

‡ Significant at the 99 per cent confidence level.

§ Significant at the 90 per cent confidence level.

‖ Significant at the 95 per cent confidence level.

that excluded education suggested that multicollinearity reduces the reliability of the estimates. At the same time, however, it is clear that excluding education results in the income coefficients being negatively biased.

With respect to the differential effects of various components of income, our analysis shows that earned income has a negative impact on middle-aged cohorts, but a positive impact on elderly cohorts. The reverse seems to hold true for unearned income. This finding, with respect to the effects of earned and unearned income, is inconsistent with Silver's study.[3] A plausible explanation of this inconsistency is that Silver's results probably are dominated by results for the elderly, since changes in mortality for these cohorts have a major impact on age-adjusted mortality. However, the results of the pure income transfer variables, WI and OUI, tend to reinforce and clarify Silver's conclusion that higher unearned income contributes to lower mortality rates. Thus, the somewhat ambiguous results for total family income can be understood better when income is broken down into earned and pure transfer components. The impact of the former is associated with age, while the latter generally has a negative association with mortality rates for all cohorts.

Another indicator of the effect of income on mortality is its elasticity, that is, the percentage change in mortality for a small change in income. Table 9 reports elasticities (evaluated at the means) for TFI and OUI, the two variables that most consistently had negative coefficients. The evidence gathered from this table is that income has a definite negative impact on mortality, but the strength of the relationship is rather weak. For both income variables considered, a 1 per cent change in income reduces mortality by about 0.05 per cent on the average. The effect is largest for the infant cohorts, with "family income elasticities" ranging from −0.047 to −0.222. On the other hand, the elasticity of OUI is larger than that for TFI for five of eight adult cohorts.

TABLE 9. Income Elasticities, All Cohorts

Cohorts	TFI[*]	OUI[*]
Adults		
WM, 45–64	−0.071	−0.017
WM, 65+	−0.020	−0.049
BM, 45–64	−0.116	−0.005
BM, 65+	0.012	0.011
WF, 45–64	−0.031	−0.039
WF, 65+	0.000	−0.035
BF, 45–64	0.019	−0.013
BF, 65+	−0.011	−0.023
Male Infants		
Whites	−0.187	−0.014
Blacks	−0.047	−0.004
Female Infants		
Whites	−0.222	−0.012
Blacks	−0.157	0.006

[*] Based on coefficients of Total Family Income (TFI) and Other Unearned Income (OUI) from Tables 5 and 6, and mean values of TFI and OUI from Table 3.

From a policy viewpoint, these results indicate that income transfer policies aimed at reducing mortality rates generally will contribute to lower mortality, but the magnitude of the effect will not be large. Since reducing mortality rates generally is not the primary motive for income transfers, this conclusion should not be interpreted as a justification for cutting back income transfer programs. Rather, this result reinforces the argument that transfer programs have payoffs in addition to the financing of current consumption.

References

1. American Medical Association. Report of the commission on the cost of medical care, vol 3. Chicago: American Medical Association, 1963-64.

2. Auster R, Levenson I, Sarachek D. The production of health, an exploratory study. J Hum Resour 1969;4:412.

3. Silver M. An economic analysis of spatial variations in mortality rates by age and sex. In: Fuchs VR, ed. Essays in the economics of health and medical care. New York: National Bureau of Economic Research, 1972.

4. Altenderfer ME. Relationship between per capita income and mortality in the cities of 100,000 or more population. Public Health Rep 1947;62:1681.

5. Antonovsky A. Social class, life expectancy and overall mortality. Milbank Mem Fund Q 1967;48:31.

6. Lerner M, Stutz RN. Mortality by socioeconomic status, 1959-61 and 1969-71. Md State Med J 1978.

7. Kitigawa EM, Hauser PM. Differential mortality in the United States. Cambridge: Harvard University Press, 1973.

8. Fuchs VR. The contribution of health services to the American economy. Milbank Mem Fund Q 1966;44:65.

9. Grossman M. The demand for health: a theoretical and empirical investigation. New York: National Bureau of Economic Research, 1972.

10. Orcutt GH, Franklin SR, Mendelsohn R, Smith JR. Does your probability of death depend on your environment? A microanalytic study. American Economic Review 1977;67:260.

11. Reid M. An exploration of conflicting evidence of death rates and income. Unpublished data, Department of Economics, University of Chicago, 1976.

12. U.S. Department of Health and Human Services, National Center for Health Statistics. Geographic patterns in the risk of dying and associated factors ages 35-74 years, United States, 1968-72. Hyattsville, Md.: USDHHS, NCHS, 1980.

13. Hadley J. Assessing the adequacy of health manpower supply. Urban Institute Working Paper 1225-5, Washington, D.C., November 1980.

14. Sheps MC. Marriage and mortality. Am J Public Health 1961;51:547.

15. Johnston J. Econometric methods. New York: McGraw-Hill, 1972.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Income and Race Differences in Children's Health in the Mid-1960s

LINDA N. EDWARDS, PH.D.,* AND MICHAEL GROSSMAN, PH.D.†

This article explores income and race differences in eight measures of the health of children ages 6 through 11 as assessed in the early 1960s. It is shown that both income and race differences in health are much less pronounced than they are in infant mortality and birth weight data. Significant differences are found in the health status of black and white children and of children from high- and low-income families, but these are primarily differences with respect to parent-reported (rather than physician-reported) health criteria and they by no means overwhelmingly favor the white or high-income children. These findings underscore the importance of treating children's health status as multidimensional. In addition, these findings will serve as a bench mark for studies of children's health using data for a more recent period.

RECENT REVIEWS of children's health in the United States have stressed the differences between black and white children and between children in high- and low-income families.[1,2] These studies cite mortality rates for both infants and older children that are over 50 per cent higher for blacks than for whites. They also point out that low birth weight (a good negative indicator of whether an infant will survive his first year and of his successful future development) is more prevalent among black and poor families.

The income and race comparisons in children's health status cited above, as well as those cited elsewhere, rely primarily on measures that relate to the first year of the child's life. This results largely from the unavailability of comprehensive measures of morbidity for older children. Publication of data from Cycles II and III of the Health Examination Survey, however, makes it possible to study explicitly the health of older children. In this paper we use data from Cycle II, which covers children aged 6 through 11 years in the 1963-65 period, to explore income and race differences in eight measures of children's health.

The Cycle II children were examined in the mid-1960s. Nevertheless, their records are still an important resource for analysis. These children comprise the only large representative sample of U.S. children taken from the recent past for which such a rich array of health and family background data are available. Undoubtedly, there have been changes in the health status of children in this age-group during the intervening 15 years, especially as a result of goverament health programs targeted at the poor, so that the relative status of black and poor children may have changed. But

---

* Department of Economics, Queens College and Graduate School, of the City University of New York and National Bureau of Economic Research, New York, New York.

† Department of Economics Graduate School, of the City University of New York and National Bureau of Economic Research, New York, New York.

This paper has not undergone the review accorded official National Bureau of Economic Research publications; in particular, it has not yet been submitted for approval by the Board of Directors.

Supported by PHS Grant No. 1 R01 HS 02917 from the National Center for Health Services Research to the National Bureau of Economic Research and a grant from the Robert Wood Johnson Foundation to the NBER.

Reprint requests: Michael Grossman, National Bureau of Economic Research, 269 Mercer Street, 8th Floor, New York, New York 10003.

0025-7079/82/0900/0915/$01.60 © J. B. Lippincott Co.

even so, analysis of the data from 1963 to 1965 will provide a benchmark against which to measure that change as future samples become available.

We report two kinds of results. First, we show that when health measures from mid-childhood are the subject of analysis, both income and race differences are much less pronounced than they are in infant mortality data. Indeed, for one of our most comprehensive health measures—the examining doctor's assessment of whether or not the child has any "significant acquired abnormalities"—there are no statistically significant differences either by income or by race. Second, the differences that we do find do not uniformly favor children from white or higher-income families.

Measuring Health in the Cycle II Sample

Cycle II of the Health Examination Survey, conducted by the National Center for Health Statistics (NCHS), is an exceptional source of information about a national sample of 7,119 noninstitutionalized children aged 6 to 11 years in the 1963-65 period. The data comprise medical and developmental histories of each child provided by the parent, information on family socioeconomic characteristics, birth certificate information and a school report with data on school performance and classroom behavior provided by teachers or other school officials. Most important, there are objective measures of health from physical examinations administered by the Public Health Service. Together, the data provide an unusually detailed picture of the health of children in this cohort.

A full description of the sample, the sampling technique and the data collection is presented in NCHS.[3] An exceptional strength of this sample is its high response rate of 96 per cent with nonwhites having a slightly higher response rate than did whites (0.982 versus 0.956). (See NCHS,[3] Table 4.) The one deficiency of the sample, from the point of view of studying children's health, is the exclusion of children in institutions. To the extent that these children are more likely to have serious and disabling physical conditions, the reported incidence of certain conditions will be lower in our sample than in the entire population of children. In addition, if the probability of the institutionalization of a child with a given condition depends on income and race, our results will incorporate unknown biases. The number of institutionalized children is small, however, at about four-tenths of a per cent of all children aged 5 through 13 years. This is the proportion of 5–13 year olds living in "group quarters" in 1970 according to the U.S. Bureau of the Census,[4] Tables 52 and 205. The corresponding percentages by race are 0.38 per cent for whites and 0.7 per cent for blacks.

Given the unusual detail and diversity of the health data in the Cycle II survey, the choice of health status measures is not dictated solely by data availability, but rather is an issue that must be faced directly. The problem of defining and measuring children's (and adults') health is very much an unresolved one, even among professionals in the area of public health.[5,6] The economist's approach is to define health as a form of human capital that determines the amount of time available for consumption and for work in the home and labor market.[7] With this type of definition, an appropriate measure of health status over some time period would be the proportion of one's potential time that is actually available for the usual consumption, maintenance and work activities. Similarly, the complementary measure of ill health would be that proportion of one's potential time lost as a consequence of imperfect functioning. Such disability may be relatively easy to measure when dealing with adults who are members of the labor force (a good approximation is days lost from work because of illness), but it is not easy to measure for other adults or for children. Therefore, in economists' studies of adults' health, the occurrence of particular physical conditions and the individual's own assessment of his health status have been used as supplementary health measures.[8]

We use the same type of restricted, morbidity-oriented, definition of children's health, focusing on the child's physical health rather than his overall well-being, and similar types of measures including disability, physical conditions, and parental assessment of health status. Even some of these measures, however, may not always be appropriate for children, because certain childhood diseases and acute conditions that prevent children from carrying out their normal activities may not reflect their health capital or "permanent" health. A useful distinction to make here is between permanent health which is one's prospect for life preservation and normal lifetime functioning, and "transitory" health, which is short-run deviations from one's normal state of health. It is the child's permanent health status that we wish to study, and we attempt to choose health measures that are indicators of that permanent health status.

Of course, there is a positive relationship between permanent and transitory health status, in the sense that a child with low health capital is more likely to contract some acute conditions and to have them for a more extended time period. For example, Birch and Gussow[9] discuss how a child's nutrition (which is clearly a determinant of permanent health status) and his likelihood of sustaining an acute condition are intimately related. Similarly, Starfield[10] reports that some acute problems may have a high likelihood of permanence in some population subgroups.

In some situations a single overall index of permanent health might be desired to help one describe parsimoniously the health status of a population, for example, or allocate public funds. Infant mortality statistics frequently are used in this way. Health, however, is clearly a multidimensional concept. Consequently, we use a set of health measures rather than a single index. Analysis of a set of component measures also avoids the essentially arbitrary weighting of the various dimensions of health implied by a health index.‡ Finally, such analysis allows for the possibility that the various aspects of children's health are differentially related to family income and race.

The actual choice of components of children's health status to be examined has been guided by the child health literature,[11–19] as well as by discussions with public health specialists and pediatricians. In addition, we have tried to take advantage of the diversity of the data by including measures derived not only from the physician's examination but also from the parent's report and the school report. The latter two types of measures, of course, are subjective, in that they depend on the perceptions of the parent or teacher. The final set of measures we study, listed and described below, will not satisfy everyone. But some exclusions result from unavailability (anemia and corrected distance vision, for example) or extremely low prevalence (poor hearing and injuries due to accidents), rather than deliberate choice.

1. The presence of one or more "significant acquired abnormalities" on physical examination of the child, represented by the dichotomous variable ABNORMALITY. These abnormalities include heart disease; neurological, muscular or joint conditions; and other major diseases.[§]
2. The child's visual acuity, represented by the dichotomous variable VISION. VI-

---

[‡] In earlier work some attempts were made to condense the health information with the use of principal component analysis. The analysis yielded almost as many equally weighted components as there were initial health measures.

[§] In defining ABNORMALITY, we exclude abnormalities resulting from accidents or injuries, because these may be more likely to reflect transitory rather than permanent health variations. The proportion of children in the full sample that had abnormalities due to accidents or injuries (but no other type of abnormality) is 2 per cent. Note that congenital abnormalities are not studied, because they generally relate to events that occur prior to birth (the prenatal period is not investigated in this paper).

SION indicates whether the child has abnormal distance vision. All children were examined only without their eyeglasses; their uncorrected binocular distance vision was defined as abnormal if it was worse than 20/30.[20] It also would have been desirable to investigate corrected distance vision, but the Cycle II data do not include the information necessary to do so.

3. The child's blood pressure, represented by BLOODP. BLOODP is a dichotomous variable which indicates if the child's diastolic blood pressure exceeds the 95th percentile for his or her age and sex.

4. The child's periodontal index, represented by PERI. The periodontal index is described as follows:[21] "Every tooth in the mouth . . . is scored according to the presence or absence of manifest signs of peridontal disease. When a portion of the free gingiva is inflamed, a score of 1 is recorded. When completely circumscribed by inflammation, teeth are scored 2. Teeth with frank peridontal pockets are scored 6 when their masticatory function is unimpaired and 8 when it is impaired. The arithemetic average of all scores is the individual's [periodontal index], which ranges from a low of 0.0 (no inflammation or periodontal pockets) to a high of 8.0 (all teeth with pockets and impaired function)."[21] PERI is a good overall indicator of oral health as well as a correlate of nutrition.[22] Since the periodontal index is known to differ systematically by age and sex, PERI is standardized by age and sex (PERI is measured as the difference between the child's actual periodontal index and the average value of that index for his or her age-sex group divided by the standard deviation of the index for that age-sex group). Higher values of PERI denote poorer values of oral health.

The periodontal index, however, suffers from the defect that it is subject to intrarater and inter-rater variability. We have experimented with a somewhat more objective measure of oral health, the number of decayed permanent and primary teeth adjusted for age and sex, and have obtained results similar to those for the periodontal index. We chose to report on the periodontal index rather than on the prevalence of decay because we judged that the former is more likely to reflect underlying nutritional and health habits and less likely to simply indicate the receipt of recent dental care.

5. The parent's assessment of the child's overall current health, represented by ASSESS. ASSESS is a dichotomous variable indicating whether the parent views the child's health as either "poor," "fair" or "good" as opposed to "very good." This particular dichotomy was chosen because there were very few children rated in the "poor" and "fair" categories (5.0 per cent) whereas a large number of parents felt it pertinent to distinguish between "good" and "very good."

It is interesting to note that although this is a subjective measure, the NCHS[23] reports that it is positively related to the physician's overall assessment of the child's health. That is, children who are given a lower health rating by one of their parents are more likely than average to be rated by the physician as having a "significant abnormality."

6. Whether or not the parent reports that the child has hay fever, asthma or "other allergies," represented by the dichotomous variable ALLERGY.

7. The parent's assessment of the child's level of tension, represented by the dichotomous variable TENSION. TENSION indicates if a child is rated by his or her parent as "high strung" or "moderately tense," as opposed to "moderately relaxed" or "unusually calm and relaxed." This dichotomy, again, was partially dictated by the data; very few of the children were rated in the "high strung" category.

8. Excessive absence from school for health reasons during the past 6 months, represented by the dichotomous variable ABSENT. This variable is taken from information provided by the child's teacher or a school administrator.

As is implied in the above definitions, all of these variables are negative correlates of good health. Note that four of the variables are taken from the physical examination (ABNORMALITY, VISION, BLOODP AND PERI), three are taken from the medical history form completed by the parent (ASSESS, ALLERGY and TENSION) and one is taken from the school form (ABSENT). Note also that variables taken from the physical examination are subject to inter-rater variability in that physicians and dentists may differ in the way they interpret the results of an examination.[24,25] This means that both objective health variables (those obtained from the physical examination) and subjective health variables (those obtained from the parent) are subject to measurement errors. But measurement errors in the subjective variables are likely to be correlated with income and race (see below), while measurement errors in the objective variables are unlikely to be correlated with race and income. This follows because white and black children and children from different income classes were examined by physicians and dentists with similar qualifications.

Race and Income Differences in Health Status in the Cycle II Sample

Table 1 presents the mean levels of these eight health measures for our entire working sample, as well as by family income and by race. The two family income classes are under $5,000 per annum and $5,000 or more per annum. This income cutoff is selected because it approximately identifies the lowest quartile of the income distribution for the Cycle II sample. For purposes of comparison, Table 1 also includes statistics on infant mortality and the prevalence of low birth weight for the same income and race classes.

The working sample used in Table 1, as well as in all subsequent calculations, includes only 4,777 of the 7,119 observations, because, first, we restricted our analysis to children who either lived with both of their parents or with their mothers only (no stepparents, foster parents, grandparents, etc.). We did so to guarantee that the family background variables used in the next section reflect the actual environment of the child.§ We also excluded those children for whom there were missing data for any of the health measures or socioeconomic variables used in the next section.‖ A comparison of the means of the eight health measures computed from our working sample with corresponding means for the entire Cycle II sample reveals very little difference between them in most cases (Table 2).¶ Table 2 also reveals that the entire Cycle II sample contains slightly higher proportions of black children and children from low-income families than the working sample. These differences are not, however, large enough to suggest that the results of our analysis would be altered if the entire sample had been utilized.

Looking now at Table 1, we present not only the race and income differences in the health measures, but also the standard normal statistics required for testing whether or not observed race and income differences could have arisen by chance.# These "z" statistics are computed in two ways. The first way assumes that the Cycle II data were generated by a simple random sampling procedure, while the second makes an adjustment to allow for the actual multi-stage cluster sampling design used to collect the Cycle II data.

Computing sample standard errors for this type of sample is complicated and requires information about the location of each sample observation. At the time of

---

§ Only 1 per cent of the children lived with their fathers only. The 72 children who turned 12 years old before the completion of the survey were also excluded.

‖ There is no school form for approximately 500 children in the Cycle II data set. Since excessive absence due to illness is the only variable taken from the school form, children without the school form are eliminated from our working sample only when school absenteeism is examined.

¶ All means and standard deviations discussed here and in the rest of the paper were computed from the unweighted data. Follow-up calculations with the weighted data altered the means only slightly (typically in the third decimal place only).

# For the dichotomous health measure the appropriate statistical test uses the binomial distribution. With samples as large as this one, however, the binomial distribution is well approximated by the normal distribution. For the continuous health measures, the "t" distribution is needed. But again, for large samples, the "t" distribution is well approximated by the normal distribution.

TABLE 1. Mean Values of Health Measures* in Cycle II Sample by Family Income and Race, and Infant Mortality and Birth Weight by Family Income and Race

Cycle II Health Measure	Cycle II Working Sample (n = 4777)	Blacks (n = 581)	Whites (n = 4196)	"z" Statistic for Black-White Difference	Adjusted "z" Statistic for Black-White Difference	Income < \$5,000/yr (n = 1645)	Income ≥ \$5,000/yr (n = 3132)	"z" Statistic for Low-High Income Difference	Adjusted "z" Statistic for Low-High Income Difference
ASSESS	0.4706	0.6127	0.4509	7.36	4.09	0.5939	0.4058	12.58	6.99
VISION	0.1160	0.1136	0.1163	0.20	0.11	0.0954	0.1268	3.21	1.78
BLOODP	0.0557	0.0671	0.0541	1.28	0.71	0.0614	0.0527	1.25	0.87
ALLERGY	0.1478	0.0878	0.1561	4.35	2.42	0.1015	0.1721	6.56	3.65
TENSION	0.4593	0.3373	0.4762	6.32	3.51	0.4304	0.4745	2.90	1.61
ABNORMALITY	0.0373	0.0430	0.0365	0.78	0.43	0.0389	0.0364	0.44	0.24
PERI†	−0.0392	−0.0736	0.0344	1.19	0.66	0.0711	−0.0971	7.48	4.16
ABSENT‡	0.0445	0.0422	0.0449	0.28	0.16	0.0483	0.0426	0.87	0.48
Other Health Measures	Total Population 1963-65	Black Population 1963-65	White Population 1963-65	Population With Income < \$5,000/yr. 1963–65	Populaltion With Income ≥ \$5,000/yr. 1963-65				
Infant mortality§	23.0	39.5	20.8	28.4	18.1				
Per cent of births with weight < 2,500 grams	7.87	14.01	7.01	8.89	7.05				

* Health measures defined in text.

† The mean of this variable is not zero, because standardization was done using the means and standard deviations for the entire Cycle II sample rather than for our working subsample.

‡ Means derived from subsample of working sample for which a school report was available (n = 4,333 rather than 4,777)

§ Deaths per 1,000 live legitimate births.[26]

this analysis such information was not provided by the Health Examination Survey (HES), in order to preserve the confidentiality of the Cycle II data. We obtain an approximation of the true standard errors for the Cycle II sample by adapting the results of a Monte Carlo study of data from the Cycle I HES survey (which utilizes the same sample design as Cycle II) reported by McCarthy.[27] McCarthy finds that the true standard error of sample means in the Cycle I data is, on the average, 1.8 times the value computed under the assumption that the data came from a simple random sample. He also estimates that the true standard error of regression coefficients computed from the Cycle I data is, on the average, 1.35 times the standard error computed under the assumption of simple random sampling. These two adjustment factors are applied to the estimates in an effort to control for the effects of the Cycle II sample design.

Statistically significant race differences (at the 5 per cent level, which is used throughout this paper) are reported for only three of the eight health measures: ASSESS, ALLERGY and TENSION. Black parents assess their children to be in poorer overall current health than do white parents. At the same time, black parents are less likely than white parents to report that their children are extremely tense or are afflicted with allergies. These three health measures are all subjective, since they are based on the parents' report. For the more objective measures, no race differences are statistically significant. The reported differences do suggest, however, that the black children in the sample are more likely to have physical abnormalities and high blood pressure than the white children. The data also suggest that black children are less likely to have periodontal disease and poor vision. In addition, black children are less likely than white children to be absent from school because of illness.

TABLE 2. Percentage of Black Children, Percentage of Children from Low-Income (< $5,000/yr) Families, and Mean Values of Health Measures in Cycle II Total and Working Samples*

Cycle II Measure	Total Sample	Working Sample
Percentage of black children	13.9	12.2
Percentage of children from low-income families	37.1	34.4
Health Measures		
ASSESS	0.4790	0.4706
VISION	0.0950	0.1160
BLOODP	0.0510	0.0557
ALLERGY	0.1580	0.1478
TENSION	0.4560	0.4593
ABNORMALITY	0.0420	0.0373
PERI	0.0000	−0.0392
ABSENT	0.0440	0.0445

* In computing the percentage or mean of a given measure in the total sample, observations with missing data for that measure alone are deleted.

To summarize, race comparisons of children's health based on these eight measures clearly yield a much less uniform impression than is drawn from inspection of data on infant mortality and low birth weight. Rather than exhibiting the dramatically large health deficits of black infants, older black children do not appear to have a consistently different health profile than their white counterparts. Moreover, those differences that are statistically significant only pertain to subjective measures. This divergence between the relative health profile of blacks in infancy and in midchildhood is not simply a result of differences in data sources or differences between sample data and population data. Data for the Cycle II children regarding their infancy (birth weight, the incidence of congenital abnormalities and the parents' retrospective assessment of the infant's health) indicate that even the black children in the Cycle II sample had significantly poorer health in infancy than did the white children in the sample.

Income differences in these health measures are of the same order of magnitude as are the race differences, only three being statistically significant.**[28] Two of these place children from higher income families in better health (the parents' overall assessment and the periodontal index), while the allergies measure places them in worse health. These income-related differences are independent of race and are evident in the white sample as well (Table 3). It is notable that for the doctor's overall assessment of significant abnormalities there is no significant income difference.

Data for the period 1971-72 from Cycle I of the Health and Nutrition Examination Survey, however, give a somewhat different result regarding income differences in visual acuity when children are examined with their "usual correction"[29] In particular, a smaller proportion of low-income children had usual distance vision in the better eye of 20/20 or better, compared with high-income children; and a larger proportion of low-income children had usual distance vision of 20/50 or worse. It should be noted, however, that these relationships are not uniform over the three income classes studied. A larger proportion of low-income (less than $5,000) children had vision of 20/20 or better, compared with middle-income ($5,000–9,999) children. A larger proportion of high-income ($10,000 and over) children had vision of 20/50 or worse, compared with low-income children. Moreover, tabulations for the classification 20/30 or worse are not shown.

What conclusions can be drawn from this simple comparison of means between the black–white and low-income/high-income subsamples? First, neither race nor income differences are pronounced. Second, those variables for which there are significant race and income differences are, for the most part, subjective measures. Finally, and most notably, we report that for the overall evaluation of the child's health by the Public Health Service physician (ABNORMALITY), there are no significant differences either by race or by income class.

Our results with respect to race and income differences in the parent's assessment of the child's overall health are similar to those reported by the NCHS,[23] in spite of our use of a smaller sample. Our results, with respect to significant acquired abnormalities, differ from those of the NCHS. Both analyses indicate fewer abnormalities in white children, compared with black children, and in children from high-income families, compared with children from low-income families. But the NCHS differentials are statistically significant, while those in Table 1 are not, probably because we exclude congenital abnormalities, while the NCHS includes them. The black–white differential in the proportion of congenital abnormalities is significant in our sample.

Decomposition of Observed Race and Income Differences

We have documented that race and income differences in health status in the Cycle II sample children are much less sharp that the corresponding differences in measures of infant health. Nevertheless, some differences still do exist. To what ex-

---

** Note that the differences in the prevalence rates of tension and abnormal uncorrected distance vision between the high-income and low-income samples are significant if the unadjusted "z" statistics in Table 1 are used. Kish[28] shows that the procedure used to adjust for design effects overstates the true standard error of the difference in means. This is because "... two subclass means from clustered samples, although based on distinct sets of elements, tend to come from the same set of clusters. The positive correlation between cluster influences on the two means tends to reduce the variance of the difference (p. 582)." Kish also shows that the estimate of the standard error of the difference in means based on the assumption of simple random sampling understates the true standard error. Hence, caution is required with respect to conclusions concerning the statistical significance of gross income difference in the prevalence of tension and abnormal vision. Note that the preceding point refers only to computing the standard error of the difference in means and is not relevant to our procedure for computing adjusted "z" statistics of regression coefficients.

TABLE 3. Gross and Net Income Differences in the Health Status of White Children*

Health Measure	High Income Mean	Low Income Mean	Gross Income Difference†	"z" Statistic on Gross Difference	Net Income Difference‡	Ratio of Net to Gross Income Difference
ASSESS	0.394	0.589	−0.195§	11.62	−0.079	0.41
VISION	0.127	0.091	0.036	3.25	0.006	0.17
BLOODP	0.052	0.056	−0.004	0.84	0.006	−1.50
ALLERGY	0.176	0.108	0.068§	5.59	0.016	0.24
TENSION	0.484	0.456	0.028	1.65	−0.002	−0.07
ABNORMALITY	0.036	0.037	−0.001	0.04	−0.001	1.00
PERI	−0.097	0.117	−0.214§	8.38	−0.077	0.36
ABSENT	0.043	0.049	−0.006	0.85	0.001	−0.17

* There are 2,718 white children in the high-income (≥ $5,000/yr) sample and 1,227 white children in the low-income (< $5,000/yr) sample.

† This is computed as the mean in the high-income sample minus the mean in the low-income sample.

‡ Computation described in text.

§ Significant income differences at the 5 per cent level of significance, whether or not adjustment is made for sample design.

tent are these uniquely associated with income and race, and to what extent can they be attributed to correlated socioeconomic factors? This question is especially relevant here because it is the subjective measures for which significant differences are observed, and these measures are most likely to be affected by systematic variations in reporting and/or scaling related to socioeconomic characteristics.

Race Differences

It is well known that race and income are highly correlated. Thus it is not surprising that two of three of the health measures displaying significant race differences also exhibit significant income differences (Table 1). An obvious first step, therefore, is to try to determine whether these observed race differences really are just a result of differences in income. To do so we obtain mean values for the eight health measures when the Cycle II sample is cross-classified by both income and race (Table 4). There are still significant race differences in health, but different measures are affected in the two income classes. Among low-income families significant race differences are observed for tension and the periodontal index, and for both of these measures black children are rated in better health than white children from families of comparable low-income levels. In high income families, significant race differences are reported for the parents' assessment, allergies and tension, black children being rated in worse health according to the parents' assessment and in better health according to tension and allergies. Thus, within income classes significant health differences still exist between black and white children. As before, these differences do not uniformly favor children of either race, but most of the significant differences are for the subjective measures and most of them show blacks to have a better rating.

These results can again be contrasted with comparable data for the two infant health measures discussed earlier. Even within income classes, race differences in infant mortality and in the incidence of low birth weight remain large and consistently favor whites (Table 5).

TABLE 4. Gross and Net Race Differences in Children's Health Status

Health Measure	Gross Difference (Black Mean-White Mean)	Low-Income Families (< $5,000/yr)				High-Income Families (> $5,000/yr)				Net Difference
		Black Mean (n = 418)	White Mean (n = 1,227)	Difference	"z" Statistic for Black-White Difference	Black Mean (n = 163)	White Mean (n = 2,969)	Difference	"z" Statistic for Black-White Difference	
ASSESS	0.162*	0.612	0.589	0.023	0.89	0.614	0.394	0.220*	5.57	0.119†
VISION	−0.002	0.108	0.091	0.017	0.98	0.129	0.127	0.002	0.10	0.009‡
BLOODP	0.013	0.069	0.056	0.013	0.79	0.061	0.052	0.009	0.51	−0.003‡
ALLERGY	−0.068*	0.084	0.108	−0.024	1.39	0.098	0.176	−0.078*	2.57	0.205†
TENSION	−0.139*	0.354	0.456	−0.102*	3.66	0.295	0.484	−0.189*	4.74	−0.153*‡
ABNORMALITY	0.007	0.046	0.037	0.009	0.80	0.037	0.036	0.001	0.03	0.018‡
PERI	−0.039	−0.063	0.117	−0.180*	3.63	−0.101	−0.097	−0.004	0.10	−0.059‡
ABSENT	−0.003	0.045	0.949	−0.004	0.32	0.034	0.043	−0.009	0.51	−0.005‡

* Significant race differences in means at the 5 per cent level of significance. These results hold whether or not an adjustment for the sample design is made in the variances with the exception of ALLERGY in the high-income families, in which case the adjusted "z" is 1.43.

† Net difference equals difference in health levels predicted for blacks and whites assuming that all explanatory variables are fixed at their mean values for the full working sample. We do not present the results of significance tests in these cases because of computational difficulty.

‡ Net difference equals regression coefficient of a race dummy variable (1 = black) from a pooled regression of black and white children that holds constant all other independent variables.

A further way of investigating the nature of race differences in children's health is to look at residual race differences after a much larger list of socioeconomic variables is held constant. This is done with the use of multiple regression analysis.†† The dependent variables in the regression equations are the eight health status measures. For the explanatory variables a set suggested by the economic model of family investment in children's health described in Edwards and Grossman was used.[30] Included are family income; parents' educational attainment; whether the child's father lives with the family; whether the child is a twin or a first-born; whether a foreign language is spoken in the home, an indication of the region of residence and size of city of residence; and the sex of the child (the latter is included only for health measures that are not standardized by sex). These variables are defined in detail in Edwards and Grossman.[30]

TABLE 5. Infant Mortality and Birth Weight

	Low Income (< $5,000/yr.)		High Income (≥ $5,000/yr.)	
	Blacks	Whites	Blacks	Whites
Infant mortality	43.9	24.2	23.6	18.6
Incidence of low birth weight	13.2	7.6	17.1	6.6

Sources and definitions of health measures are the same as in Table 1.

Race differences in children's health net of differences in this set of socioeconomic variables are presented in the last column of Table 4. Net differences are computed in two different ways, depending on whether or not there are significant race differences in slope coefficients in the underlying health equations. If there are no significant differences in the slope coefficients by race (as in the case for the dependent variables, BLOODP, VISION, TENSION, ABNORMALITY and ABSENT), the net race difference is represented by the regression coefficient of a race dummy variable (black = 1) from a pooled regression of black and white children that holds constant all other dependent variables. When there are significant race differences in the slope coefficients (as in the case of the dependent variables PERI, ASSESS and ALLERGY), the net difference is computed according to the following formula: Net difference = $(a_b + b_b \overline{X}) - (a_w + b_w \overline{X})$, where a_b and b_b represent the intercept and vector of slope coefficients estimated when one is using the black subsample, a_w and b_w represent the corresponding coefficients estimated when one is using the white subsample, and $\overline{X}$ represents a vector of means of the set of explanatory socioeconomic variables computed over the entire working sample. In either case, the net difference can be interpreted as that portion of the gross difference that would remain if both race groups had the same mean values of all explanatory variables.

In general, the net differences are not smaller than the gross differences. Indeed, in some cases they are substantially larger and /or change signs. For example, for the tension variable, the net difference is greater than the gross difference, implying that if both black and white families had the same socioeconomic characteristics, the blacks relatively would be even less likely than whites to report excessive tension. Or, in the case of allergies, the net difference implies that blacks would be more likely than whites to report allergies if both blacks and whites had the same socioeconomic characteristics. The net difference for the ABNORMALITY variable is also greater than the gross difference, but it still does not increase enough to reach statistical significance. The one important exception to the generalization above is the parents' assessment variable. In this case, the net difference is less than the gross difference.

---

†† Ordinary least squares was used to estimate linear equations using the health measures as dependent variables. When the health measures are dichotomous, this procedure is not correct because the residuals variance will exhibit heteroscedasticity. However, experiments with a more appropriate (and much more costly) estimation procedure—LOGIT—indicated that our results are virtually unchanged when LOGIT is used in place of ordinary least squares.

To sum up, the various statistics in Table 4 suggest the following conclusions. First, our view of race differences in children's health is not altered substantially when income and other socioeconomic variables are held constant. Second, these differences do not obviously favor children of either race. Third, it is notable that for some of the harder measures—the physician's overall assessment, unusually high blood pressure, poor vision—significant race differences are never reported. On the other hand, differences in subjective measures tend to be significant. The latter may reflect actual differences or it may reflect only perceived differences. It may even reflect systematic differences in scaling: for example, a family undergoing many stresses (as a low-income family might) may not place much weight on a child's symptoms of allergy or tension. Indeed, it may only be a visit to a doctor that makes a family aware that a child is afflicted with allergies or is excessively tense. Thus, it is clear that results for these subjective measures must be interpreted with extreme caution.

Income Differences

Gross race differences in children's health did not disappear when various socioeconomic factors (including income) were held constant. Does the same conclusion hold for income-related differences? In other words, to what extent do the gross differences in health status between income classes disappear when additional socioeconomic factors are held constant? To answer this question, we use the same type of multiple regression analysis described earlier. We simplify the analysis, however, by restricting the decomposition of gross income differences to the white sample only.

According to the adjusted "z" statistic, the list of health measures for which there are significant income differences is the same for the white sample as it is for the full black-and-white sample. If the unadjusted "z" statistic is used, however, there is no significant income difference for tension in the white sample, but there is in the white–black sample.

Gross and net income differences for the eight health measures are presented in Table 3. The net income differences are computed similarly to the corresponding net race differences. They should be interpreted as the difference in mean health status between the two income classes if all of the socioeconomic variables (other than income, of course) took the same values in both classes. Unlike the case of race differences, there were no significant differences for health measures in the slope coefficients in the two income classes. The calculations, however, do permit the coefficient of income to be different among high- and low-income families. The theoretical rationale for this is discussed in Edwards and Grossman.[30] The net income difference for any health measure is computed as the coefficient of income multiplied by the difference in mean income in the two income classes. Note that in these equations income is measured as a continuous variable, whereas race is dichotomous. Income in Cycle II is reported in ten classes: less than \$500, \$500–\$999, \$1,000–\$1,999, \$2,000–\$2,999 \$3,000–\$3,999, \$4,000–\$4,999, \$5,000–\$5,999, \$6,000–\$6,999, \$7,000–\$9,999, \$10,000–\$14,999, and \$15,000 and over. We convert these classes into a continuous variable by assigning midpoints to the closed intervals, \$250 to the lowest interval and \$20,000 to the highest interval.

The number of health status measures for which statistically significant income differences exist is reduced from three (ASSESS, ALLERGY and PERI) to zero when related socioeconomic factors are held constant. For these three (and for most of the other measures as well) the magnitude of the net difference is substantially smaller than the corresponding gross difference (see the last two columns of Table 3). For example, on the basis of the gross difference, about 20 per cent more of the high-income parents than of the low-income parents assessed their children's health as being very good. When related socioeconomic factors are held constant, this differential is reduced to only 8 per cent. To conclude, no significant income differences in health status are reported when related socioeconomic characteristics are held constant, and the magnitudes of all of these income differences are greatly diminished. The main conclusion to be drawn is clear: gross income differences in health greatly overstate the true relationship between family income and health.

Further insight into the nature of gross income differentials in health status is obtained by study of the precise role of the explanatory socioeconomic variables. Table 6 presents calculations which illustrate how the gross income differences are decomposed among the various explanatory factors for the three health variables that exhibited significant gross income differences in Table 3. The procedure simply is to multiply the coefficients of these explanatory variables by the differences in their mean values in the high- and low-income samples of children. Several results in Table 6 are noteworthy. First, more than 80 per cent of the differences in the four health measures between the high- and low-income subsamples can be accounted for by differences in the independent variables that we have included in our multiple regression equation. Second, a detailed examination of the decomposition indicates that differences in parents' average schooling between high- and low-income families account for about as much or even more of the gross differences as does income itself. And finally, more of the income-related gross differences in children's health are accounted for by differences in correlated socioeconomic and regional characteristics of the child than by family income itself.

Our findings with respect to race and income differences may be compared with those in a recent study of infant mortality

TABLE 6. Components of the Difference in ASSESS, ALLERGY and PERI Between White Children from High- and Low-Income Families*

Component	ASSESS	ALLERGY	PERI
Family income	−0.079	0.016	−0.077
Parents schooling	−0.071	0.045	−0.131
Other family characteristics †	0.001	0.004	−0.011
Characteristics of child ‡	−0.003	0.003	0.001
Region	−0.025	−0.005	0.019
City size	−0.005	−0.004	−0.001
Total = predicted gross difference	−0.182	0.059	−0.200
Actual gross difference	−0.195	0.068	−0.214

[*] Computed from coefficients in a multiple regression which includes the explanatory variables listed in the text.

[†] These include whether or not a foreign language is spoken in the home and whether or not the father is absent from the home.

[‡] These include whether or not the child is a first-born or a twin, and his or her sex. The latter is not included for PERI which is standardized by sex.

by Gortmaker.[31] He attempts to determine what portion of the large income and race differences in infant mortality can be explained by differences in parents' educational attainment, mother's age, the child's birth order and the previous pregnancy experience of the mother. He finds that poor families and black families still display a much higher incidence of infant mortality even when these factors are controlled for. This contrasts with the findings in this article for later childhood. Once again, we are struck by the different views of child health from the alternative perspectives of infancy and midchildhood.

Conclusions

The first point that must be made in concluding is that income and race differences in infant mortality provide a poor and even misleading description of income and race differences in the health of older children for the 1963-65 period. This suggests that while it may be appropriate for one to use infant mortality statistics in broad across-country comparisons of the health status of various populations, one should not use these statistics to indicate the relative health statuses of various groups of older children within the United States.

We do find differences in the health status of black and white children and of children from high- and low-income families, but these differences by no means overwhelmingly favor the white or high-income children. With respect to differences by race, whether or not they are adjusted for differences in associated socioeconomic factors, significant differences exist primarily for the subjective health measures, and these do not always show black children to be rated in poorer health than their white counterparts. In the case of income differences in health, the high-income children do appear to be rated healthier according to most measures, but their advantage is diminished greatly when one controls for related socioeconomic factors like parents' educational attainment. It is important that one notes, however, that with respect to the variable that might reflect the most serious health problems—whether or not the doctor reports the child to have a "significant abnormality"—there are never statistically significant differences either by race or by income.

These results are especially intriguing when one considers that the children in the Cycle II sample were growing up in the late 1950s and early 1960s, a period prior to the introduction of a number of government health programs targeted at the poor such as Medicaid, maternal and infant care projects, and community health centers. A natural follow-up to our study would be similar analyses using a more recent national sample of children.

Implicit in our discussion is the necessity for recognizing the multidimensional nature of health. Our results clearly differ for different health status measures. Poor and black children tend to be rated in better health when subjective health measures are studied, but they are rated as being in the same or worse health when more objective health measures are used. Thus, our findings underscore the importance of treating children's health status as multidimensional, and illustrates how the use of a single health index could lead to erroneous conclusions about health status and its relation to income and race. Of course, when a set of measures is used instead of a single measure, the researcher is faced with more difficulty in synthesizing his results. But the fact that one is unlikely to be able to draw neat conclusions from a multifaceted analysis of health is not an acceptable reason to restrict that analysis to a single health measure.

The primary purpose of this article has been to expand upon commonly held notions about income and race differences in children's health by the use of a study of children who were 6 to 11 years old in 1963-65. We have found that the unambiguous picture drawn from infant mortality statistics becomes fuzzy and difficult to decipher when a collection of morbidity measures provides the medium. This raises an important and, as yet, unanswered question, especially if our findings are reinforced by other studies using different measures and different time periods. Why are the income and race differences in infant mortality so striking when corresponding differences in mid-childhood morbidity are not? This divergence is not simply a result of the type of measure (mortality vs. morbidity); differences in the mortality of older children also are less striking. Perhaps the health differences between various demographic groups are minimized in the mid-childhood years. If so, we need to know why. The answer to these questions are clearly pertinent to the conduct of public policy toward the welfare of children.

Acknowledgments

The following physicians gave us helpful advice: John McNamara, M.D., then Assistant Professor of Public Health and Pediatrics at Columbia University School of Public Health and Associate Commissioner in the New York City Department of Health; Roy Brown, M.D., Associate Professor of Community Medicine and Pediatrics at the Mount Sinai School of Medicine of the City University of New York; Thomas Travers, D.D.S., Director of Ambulatory Care in the New York City Department of Health; and Ruth T. Gross, M.D., Professor of Pediatrics and Director of Ambulatory Pediatrics, Stanford University Medical Center, Stanford, California. We are also indebted to Mary Grace Kovar for her helpful comments on an earlier draft.

References

1. Newberger E, Newberger CM, Richmond JB. Child health in America: toward a rational public policy. Milbank Mem Fund Q 1976;54:249

2. Keniston K. All our children: The American family under pressure. New York: Harcourt Brace Jovanovich, 1977.

3. National Center for Health Statistics. Plan, operation, and response results of a program of children's examination. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 1, No. 5, 1967.

4. U.S. Bureau of the Census. 1970 census of population, Vol I, characteristics of the population, Part 1, U.S. summary, sections 1 and 2. Washington, D.C.: U.S. Government Printing Office, 1973.

5. Sullivan DF. Conceptual problems in developing an index of health. U.S. Department of Health, Education and Welfare, Public Health Service, Series 2, No. 17, 1966.

6. Berg RL. Health status indexes. Chicago: Hospital Research and Educational Trust, 1973.

7. Grossman M. The demand for health: a theoretical and empirical investigation. New York: Columbia University for the National Bureau of Economic Research, 1972.

8. Grossman M. The correlation between health and schooling. In: Terleckyj NE, ed. Household production and consumption. New York: Columbia University Press for the National Bureau of Economic Research, 1975:147.

9. Birch HG, Gussow JD. Disadvantaged children: health, nutrition, and school failure. New York: Harcourt, Brace and World, 1970.

10. Starfield B. Stability and change: another view. Am J Public Health 1981;71:301.

11. Wallace HM. Health services for mothers and children. Philadelphia: WB Saunders, 1962.

12. Mechanic D. The influence of mothers on their children's health, attitudes and behavior. Pediatrics 1964;33:444.

13. Talbot NB, Kagan J, Eisenberg L, eds. Behavioral science in pediatric medicine. Philadelphia: WB Saunders, 1971.

14. Kaplan RS, Lave LB, Leinhardt S. The efficacy of a comprehensive health care project: an empirical analysis. Am J Public Health 1972;62:924.

15. Hu T. Effectiveness of child health and welfare programs: a simultaneous equations approach. Socio-Economic Planning Sciences 1973;7:705.

16. Kessner DM. Assessment of medical care for children, vol 3, Contrasts in health status. Washington, D.C.: Institute of Medicine, 1974.

17. Haggerty RJ, Roghmann KJ, Pless IB. Child health and the community. New York: John Wiley and Sons, 1975.

18. Inman RP. The family provision of children's health: an economic analysis. In: Rosett R, ed. The role of health insurance in the health services sector. New York: Neale Watson Academic Publications, 1976:215.

19. Starfield B. Health needs of children. In: Children's medical care needs and treatment. Harvard Child Health Project Report Vol II. Cambridge: Ballinger Publishing, 1977:23.

20. National Center for Health Statistics. Binocular visual acuity of children: demographic and socioeconomic characteristics. U.S. Department of Health, Education and Welfare, Public Health Ser-

vice, Vital and Health Statistics, Series 11, No. 112, 1972.

21. Kelly JE, Sanchez M. Periodontal disease and oral hygiene among children. National Center for Health Statistics, U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 11, No. 117, 1972.

22. Russell AL. A system of classification and scoring for prevalence surveys of periodontal disease. Journal of Dental Research 1956;35:350.

23. National Center for Health Statistics. Examination and health history findings among children and youths. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 11, No. 129, 1973.

24. Koran LM. The reliability of clinical methods, data and judgments, part one. N Engl J Med 1975;293:642.

25. Koran LM. The reliability of clinical methods, data and judgments, part two. N Engl J Med 1975;293:695.

26. National Center for Health Statistics. Infant mortality rates: socioeconomic factors. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 22, No. 14, 1972.

27. McCarthy PJ. Pseudoreplication: further evaluation and applications of the balanced half-sample technique. National Center for Health Statistics, U.S. Department of Health, Education, and Welfare, Public Health Service, Vital and Health Statistics, Series, No. 31, 1969.

28. Kish L. Survey sampling. New York: John Wiley and Sons, 1965.

29. National Center for Health Statistics. Monocular visual acuity of persons 4–74 years. U.S. Department of Health, Education and Welfare, Public Health Service, Vital and Health Statistics, Series 11, No. 201, 1977.

30. Edwards LN, Grossman M. Children's health and the family. In: Scheffler RM, ed. Advances in health economics and health services research. Greenwich, Conn.: JAI Press, 1981:35.

31. Gortmaker SL. Poverty and infant mortality in the United States. Am Sociol Rev 1979;44:280.

Erratum

The figures on Table 1, page 181, of "Predicting the Outcome of Primary Care," by D. D. Wright and R. L. Kane, *Medical Care*, February 1982, are incorrect. The predictive value of a positive test should be 0.785, and the predictive value of a negative test should be 0.571.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Regional Differences in Hospital Use Standardized by Reported Morbidity

MARK S. BLUMBERG, M.D.

The purpose of the study was to determine whether regional differences in hospital use parallel regional morbidity differences. It showed that they do not. We used the 1975 U.S. Health Interview Survey Metropolitan sample under age 65 to analyze the nonobstetric hospital use and the reported morbidity in the four U.S. regions. The West sample used less hospital care than those in other regions, even though other regions reported less morbidity. We indirectly standardized hospital use by age-sex specific morbidity. The standardized ratio of actual to expected hospital nights in the West was half that of the North Central Region. There were smaller regional differences in surgical hospital use than for nonsurgical hospital use. In 1975 the total U.S. metropolitan population under age 65 would have used one third fewer nonobstetric hospital nights if it had followed Western practices.

THE REGIONS of the United States differ in age-sex specific hospital use per person. In 1975 the annual number of hospital days per thousand persons was lowest in the west for each of the eight age-sex categories in Table 1. The rates in the west ranged from 57 to 67 per cent of the corresponding north central rates, which were the highest for those under age 65.

Almost all hospital care is for treatment or diagnosis of health problems. Hence the health status of a population is a key determinant of its relative need for hospital care. While age and sex commonly are used as proxies for health status, we sought more direct measures of population health by region to determine how many of the hospital use differences were attributable to health differences.

Data Source

Each year since 1958, the Bureau of the Census has conducted the U.S. Health Interview Survey (HIS) for the Public Health Service. The survey is designed and analyzed by the National Center for Health Statistics. We obtained public use computer tapes of the 1975 HIS because we were interested in the private health care coverage of respondents and that year information on this was obtained.

The total annual HIS sample of about 112,000 persons in 41,000 households is representative of the U.S. civilian-noninstitutional population. The sample interviews are staggered throughout the calendar year, thus avoiding seasonal bias. However, the ample process results in substantial understatement of both morbidity and hospital use for the total population age 65 and over. Those in older age groups are more likely than those under age 65 to be institutionalized for health

---

From the Kaiser Foundation Health Plan Inc., Oakland, California.

Reprint requests: Mark S. Blumberg, M.D., Kaiser Foundation Health Plan, Inc., One Kaiser Plaza, Oakland, CA 94612.

0025-7079/82/0900/0931/$01.50 © J. B. Lippincott Co.

problems. Older persons also have a higher mortality rate. Since the health and hospital experience of the institutional population and of decedants is not included in the Health Interview Survey, we chose to analyze the HIS sample population that was under age 65. The study was limited further to those residing in the over 200 Standard Metropolitan Statistical Areas (SMSAs) that were part of the Health Interview Survey sample.

Health Status and Hospital Use Measures

The two basic measures of morbidity from the HIS Survey are the "restricted activity day" and conditions causing a "limitation of activity." A day of restricted activity is one on which a person cuts down on his usual activities for the whole of that day because of an illness or an injury." Data on restricted activity days are collected for the two-week period prior to the interview. They include bed disability days, which in turn include hospital days. By definition, limitations of activity are caused by chronic conditions defined as those on a stipulated list or having been present for three months or more.

The data in Table 2 show that the 1975 West HIS Sample (including those not residing in SMSAs) was highest of any region in the number of restricted activity days per person per year in each age-sex category under age 65. The North Central region was frequently the region with the lowest number of restricted activity days per person. The pattern is similar for the percentage of the population with limitation of activity. These data suggest that the interregional differences in hospital use shown in Table 1 are not due to corresponding differences in morbidity by region.

The hospital use reported by the Health Interview Survey consisted of all that occurred in short-stay hospitals of any ownership or location (e.g., including federal and foreign) during the year prior to interview. While the study was limited to episodes in short-stay hospitals, it could have included some episodes in long-stay units of short-stay hospitals. To minimize the effect of such long stays we truncated the length of all sample episodes at 45 days.

TABLE 1. Non-Federal Short Stay Hospital Days* per 1,000 Population by Sex, Age and Region—U.S. National Hospital Discharge Survey[6] 1975

Sex and Age	Region				
	Northeast	North Central	South	West	All U.S.
Male					
Under 15	357	461·†	344	264°‡	365
15–44	650	789·	571	488°	634
45–64	1898	2035·	1492	1270°	1700
65+	4714	5065·	4075	3328°	4379
All	1215	1321·	1000	812°	1104
Female					
Under 15	281	366·	269	205°	287
15–44	805	1072·	830	658°	858
45–64	1840	2272·	1545	1528°	1792
65+	4385	4757·	3612	2988°	4016
All	1363	1570·	1158	946°	1280

* Excludes well newborns and obstetric deliveries.

† · = Highest value on each line.

‡ ° = Lowest value on each line.

TABLE 2. Restricted Activity Days per Person per Year by Region, Sex and Age—1975 Health Interview Survey*†

Sex and Age	Region				
	Northeast	North Central	South	West	All
Male					
Under 6	13.1	11.8	11.2° ‡	15.1$^{\bullet}$ §	12.5
6–16	10.5	9.5°	9.8	12.0$^{\bullet}$	10.2
17–24	9.9	7.7°	10.0	14.5$^{\bullet}$	10.2
25–44	12.1°	12.2	13.7	13.8$^{\bullet}$	13.0
45–64	19.2°	19.2°	24.7	25.4$^{\bullet}$	22.0
65+	29.0°	29.7	41.4$^{\bullet}$	36.0	34.3
All	14.7	13.8°	16.6	17.8$^{\bullet}$	15.6
Female					
Under 6	11.7	9.6°	10.4	16.4$^{\bullet}$	11.6
6–16	11.6	10.5	9.0°	13.8$^{\bullet}$	10.9
17–24	14.4	13.4°	13.7	16.8$^{\bullet}$	14.3
25–44	16.1°	16.1°	19.5	20.7$^{\bullet}$	18.1
45–64	23.6	21.3°	28.0	34.3$^{\bullet}$	26.2
65+	33.6°	38.7	48.2$^{\bullet}$	45.3	41.4
All	18.5	17.9°	20.8	23.6$^{\bullet}$	20.0

*Data are based on household interviews of the civilian, noninstitutional population.

† Gentile A. Disability Day U.S. 1975: National Center for Health Statistics, Data from the National Health Survey: Series 10, No. 118 (DHEW Publication No. (PHS) 78-1546).

‡ $^{\circ}$ = Lowest value on each line.

§ $^{\bullet}$ = Highest value on each line.

Obstetric deliveries are the most common cause of hospital admissions in the United States. However, hospitalization for obstetric deliveries is almost universal and hence, inclusion of obstetric deliveries could confound other hospital use rates that can be more discretionary. Furthermore, normal pregnancy is not considered morbidity by the HIS. For these reasons we excluded all hospitalization for obstetric deliveries, as well as for abortions. (By convention, the HIS source data omit all hospital care of well newborns.)

Standards for Hospital Use

We developed standards for hospital use for the 1975 HIS SMSA sample under age 65 (72,360 persons). These standards are based on hospital use, restricted activity days and limitation of activity due to chronic conditions of the U.S. SMSA sample under age 65. First the sample population was divided into those who had one or more conditions causing limitation of activity and those who did not. These two subsamples were then divided into eight age categories for males and for females, for a total of 32 cells. Finally, separate standards were derived for hospital episodes and for hospital nights.

Table 3 presents the standards used in this analysis. The value of 1.734 on the bottom line in the left column was derived as follows: there were 2,705 restricted activity days (in a 2-week period) reported by 1,002 males, age 55 to 64, who were limited in activity. This same sample reported 4,690 hospital nights in the past year. The standard is simply 4,690 nights divided by 2,705 restricted activity days, or 1.734 annual hospital nights per restricted activity day in 2 weeks.

The value of the standard for hospital episodes in this same population was based on the fact that this sample had 360 hospital episodes per 2,705 restricted activity days or 0.1331, as shown on the bottom of the second column.

TABLE 3. Annual Number of Hospital Episodes and Nights* per Restricted Activity Day (in a 2-week period:) by Age and Sex for Persons With and Without Limitation of Activity Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

	Sex			
	Male		Female	
Age	Nights	Episodes	Nights	Episodes
Persons not limited in activity				
Under 1	2.650	0.3376	2.000	0.2638
1–5	0.633	0.1605	0.546	0.1264
6–16	0.584	0.1265	0.432	0.1060
17–24	0.812	0.1600	1.772	0.1633
25–34	0.910	0.1636	1.119	0.1855
35–44	1.039	0.1728	1.511	0.2526
45–54	1.765	0.2434	1.679	0.2378
55–64	2.311	0.2582	1.472	0.1699
Persons limited in activity				
Under 1	11.882	0.5882	11.267	0.8667
1–5	2.123	0.2538	1.139	0.2025
6–16	1.096	0.1570	0.588	0.0785
17–24	1.550	0.1403	0.848	0.1168
25–34	1.429	0.1311	1.349	0.1479
35–44	2.071	0.1602	1.291	0.1238
45–54	1.693	1.1448	1.232	0.1155
55–64	1.734	0.1331	1.180	0.1126

* Excludes obstetric deliveries, abortion and well newborns.

We used these standards to derive estimates of the expected number of hospital episodes and hospital nights for each of the four U.S. regions. For example, in the Northeast region there were 226 males, age 45 to 54, limited in activity, who reported 435 restricted activity days. Their 435 of restricted activity days were multiplied by the corresponding U.S. standards (1.693 nights and 0.1448 episodes). This resulted in 736.22 expected nights and 62.97 expected episodes in their cell. The expected nights and episodes obtained in a like way for all of the 64 cells were summed for each of the four regions. These expected nights and episodes were then compared with the actual number of hospital nights and episodes in each region.

Results of Analysis

Some of the morbidity characteristics of the study samples, by region of the country, are shown at the top of Table 4. Line d shows that the percentage of the sample limited in activity varied by region with the West being the highest and the North Central region being the lowest (12.05 per cent versus 9.18 per cent). Lines e and f show that those who were limited in activity had many more restricted activity days per person per year than those who were not limited in activity. While this was true for each region, those in each category in the West had the highest number of restricted activity days per person, while those in the north cental region had the fewest number of restricted activity days per person.

Because of the greater reported morbidity in the West, this region also had the highest annual expected hospital episodes (line j) and hospital nights (line 1). Conversely, the North Central region had the lowest annual number of expected hospital episodes and nights per hundred persons. Table 5 presents the indirectly stand-

TABLE 4. Selected Morbidity and Expected Hospital* Use Measures for Persons With and Without Limitation of Activity by Region
Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

Selected Measures	Region				
	Northeast	North Central	South	West	All
Sample Persons					
a Not limited in activity	17,482	17,296	17,246	12,884	64,908
b limited in activity	2,024	1,749	1,912	1,767	7,452
c All	19,506	19 045	19,158	14,651	72,360
d Per cent of sample with limitation of activity	10.38	9.18°†	9.98	12.6 ·‡	*10.30*
Annual restricted activity days per person					
e Not limited in activity	10.10	9.99°	10.31	12.73·	10.65
f Limited in activity	56.35	55.67°	62.47	66.82·	60.24
Annual expected hospital use per 100 persons					
Episodes					
g Not limited in activity	6.80	6.74°	7.07	8.44·	7.18
h Limited in activity	28.37	27.59°	31.55	33.34·	30.18
i All	9.04	8.65°	52	11.44·	9.55
Nights					
j Not limited in activity	40.99	39.88°	42.33	50.04·	42.85
k Limited in activity	306.99	299.76°	341.52	354.54·	325.43
l All	68.59	63.75°	72.19	86.76·	71.95

* Excludes obstetric deliveries, abortions and well newborns.
† ° = Lowest value on each line.
‡ · = Highest value on each line.
§ Expected base on age, sex and morbidity.

ardized ratios for hospital episodes, nights and lengths of stay by region of the country. Each ratio in Table 5 is the observed (i.e., actual) value divided by the corresponding expected value for a particular sample population. The ratio for the total U.S. is 1.0 by definition. A ratio of over 1.0 for a sample indicates that its actual use was greater than its expected use (based on U.S. standards). A value of under 1.0 indicated that the sample's actual use as less than expected. Line e gives the indirectly standardized ratio for all hospital episodes. It is highest in the North Central region (1.175) and lowest in the West (0.786). Given the same reported age, sex and morbidity experience in each region, this indicates that if the indirectly standardized hospital episode ratio of the West had been applied to the U.S. as a whole, total U.S. hospital admission rates would have been 21 per cent lower (i.e., 100[1–0.786]). The lowest episode ratio is for males in the West not limited in activity (line a, 0.748). However, the West has the lowest ratio for lines a, b, c and d which are the most disaggregated data presented. The least regional variation in episode ratios is for females who are limited in activity (line d).

The indirectly standardized ratios for hospital nights are more varied than the ratios for hospital episodes. Thus the overall West ratio for hospital nights (line j) is only 0.64[illegible]. This indicates that if all hospital use in the U.S. had paralleled the West, total U.S. hospital nights would have been 35 per cent less than observed! The West has the lowest hospital night ratios for each

TABLE 5. Indirectly Standardized Ratios* for Hospital Episodes, Nights, And Length of Stay† for Persons With and Without Limitation of Activity by Sex and Region Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

Indirectly Standardized Ratio (Actual to Expected)	Region			
	Northeast	North Central	South	West
Episodes				
Not limited in activity				
a Male	0.911	1.200˙‡	1.115	0.748°§
b Female	1.000	1.167˙	1.059	0.768°
Limitation of activity				
c	1.229˙	1.174	0.927	0.802°
d Female	0.925	1.145˙	1.074	0.878°
All				
e Both sexes	0.980	1.175˙	1.057	0.786°
Nights				
Not limited in activity				
f Male	0.937	1.294˙	1.068	0.659°
g Female	1.126	1.186˙	1.053	0.619°
Limited in activity				
h Male	1.217	1.355˙	0.839	0.634°
i Female	1.085	1.304˙	1.007	0.670°
All				
j Both sexes	1.094	1.276˙	0.994	0.644°
Length of stay				
Not limited in activity				
k Male	1.028	1.078˙	0.958	0.881°
l Female	1.126˙	1.016	1.044	0.813°
Limited in activity				
m Male	0.990	1.154˙	0.905	0.791°
n Female	1.173˙	1.135	0.938	0.763°
All				
o Both sexes	1.164˙	1.086	0.940	0.819°

* Standardized for age, sex and morbidity.

† Excludes obstetric deliveries, abortions and well newborn.

‡ ˙ = Highest value on each line.

§ ° = Lowest value on each line.

of the disaggregated cells (lines f, g, h and i), while the North Central area has the highest ratios for these cells.

The data on line o are the indirectly standardized ratios for lengths of hospital stay by region of the country. The West has the lowest length-of-stay ratio (0.819), while the ratio in the Northeast is the highest (1.164). However, the North Central area also has a ratio above 1.

The extremely low hospital night ratio of the West (0.644) is due almost equally to a lower episode ratio (0.786) and to a lower length-of-stay ratio (0.819) for that region. In contrast, about two thirds of the high night ratio for the North Central region (1.276) is due to a high episode ratio (1.175) and about one third due to a high length-of-stay ratio (1.086).

The actual annual hospital use rates and lengths of stay for our HIS sample are given in Table 6, separately for cases which were and were not surgically treated. There was much more variation by region in these rates for cases that were not treated surgically (lines b, e and h) than for

TABLE 6. Actual Annual Hospital Episodes and Nights* per 100 Persons and Average Length of Stay Surgically vs. Non-Surgically Treated by Region Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

Hospital Use Measures	Region				
	Northeast	North Central	South	West	U.S.
Annual hospital episodes					
Per 100 persons activity					
a. Surgically treated	4.65°†	5.30	5.48·‡	5.04	5.12
b. All other	4.21	4.86·	4.58	3.95°	4.43
c. Total	8.86	10.17	10.06	9.00	9.55
Annual hospital nights					
Per 100 persons activity					
d. Surgically treated	35.30	37.91·	37.15	30.24°	35.45
e. All other	39.75	43.40·	34.63	25.60°	36.49
f. Total	75.05	81.31·	71.78	55.85°	71.95
Average length of stay (nights)					
g. Surgically treated	7.59·	7.15	6.78	6.00°	6.92
h. All other	9.43·	8.93	7.57	6.48°	8.24
i. Total	8.47·	8.00	7.13	6.21°	7.53

* Omits obstetric deliveries, abortions and well newborns.
† ° = Lowest value on each line.
‡ · = Highest value on each line.

cases were treated surgically (lines a, d and g). For example, the annual hospital nights per 100 persons per year in the West for cases that were not treated surgically was only 60 per cent of the figure for the north central region) 25.6 vs. 43.4). The annual surgical episode rate in the West is almost equal to the U.S. rate and the rate is only a little higher in the North Central region. The average stay of nonsurgical cases in the Northeast is 45 per cent higher than it is in the West.

Validity of the Data and Method

Because the findings above were dramatic and unexpected, we were concerned about the validity and applicability of the morbidity data used. The marked interregional variations of hospital use in the Health Interview Survey are confirmed by the Hospital Discharge Survey (the source of the data in Table 1, as well as statistics from the American Hospital Association and from Medicare. Each of these sources show that 1975 was not unusual in the regional variation of hospital use.

We sought information on reported morbidity by region for other years to determine whether or not there was something unusual about the regional pattern in 1975. Data from the HIS for selected years 1963 through 1978 are presented in Table 7. The top section provides the age adjusted rates of restricted activity days per person per year by region. Note that the age adjustment used the same four age groups each year, but the standards were based on the U.S. age distribution each year. Hence the total U.S. and age adjusted regional data are not strictly comparable from year to year. However, the data are valid to compare relative morbidity by region in any given year.

The West had the highest number of restricted activity days per person per year for each year, except 1973. The North Central region was more likely to have the lowest number of restricted activity days per person per year than any of the other regions. Thus 1975 was not unusual in the regions' rank order of restricted activity days per person.

The South tended to have the highest percentage of the population with limited activity on an age-adjusted basis, while the Northeast was consistently the lowest by this measure. The data on Table 7 are for the entire HIS population, whereas our study was for those under 65 residing in metropolitan areas. The nonmetropolitan population of the South has a high percentage with limitation of activity. Data for the years 1973 and 1974 combined show that the percentage of the population with limitation of activity residing in metropolitan areas was highest in the West for those under 45 and for those age 45–64.[9] Thus for those two years for which we have data comparable to ours, interregional differences in limitation of activity were similar to 1975.

We do not know whether there are important regional differences in the reporting of morbidity by the HIS for populations that are identical in their "true" morbidity. Given the fact that most in the West are either in-migrants or descended recent in-migrants from other regions of the U.S., we doubt any pervasive cultural difference.

In a recent editorial, Wilson[10] warns that greater use of health care services by respondents can lead them to report more morbidity in the Health Interview Survey. Data in Table 8 show that residents of Western SMSAs did make more visits to physicians per year for each adult age category than those living in other regions of the U.S. We interpret this as a response to the higher morbidity in the West, but Wilson's view cannot be rejected. It may well be that some of the higher physician visit rates in the West are a result of substituting office care for hospital care in that

TABLE 7. Restricted Activity Days Per Person Per Year and Per Cent of Population With Limitation of Activity Selected Years 1963–1978 by Region U.S. Civilian Non-Institutional Population Sample in Health Interview Survey

	Region				
	Northeast	North Central	South	West	U.S.
Restricted activity days per person per year (age adjusted)					
July 1963–June 1965[2]	13.7°*	16.0	17.9+†	17.9+	16.3
1968[3]	17.4	13.8°	16.0	17.8+	15.3
1969 & 1970[4]	14.1	13.0°	15.8	16.4+	14.7
1971[5]	14.6	14.2°	16.7	17.9+	15.7
1972[6]	15.4°	15.7	17.5	18.1+	16.5
1973[7]	13.6°	15.4	18.5+	18.2	16.4
1977[6]	16.8	16.7°	17.9	19.1+	17.4
1978[7]	17.3	16.8°	19.1	20.7+	18.4
Per cent with limitation of activity (age adjusted)					
July 1963–June 1965[2]	9.2°	11.7	14.8+	12.2	12.1
1969 & 1970[4]	10.7°	10.9	13.1+	12.1	11.7
1972[6]	11.4°	12.0	13.5	13.6+	12.6
1973[7]	12.1°	12.6	14.5+	14.1	13.1
1974[8]	13.2°	13.7	14.9+	14.8	14.1
1977[6]	12.0°	12.3	14.0+	13.5	13.0
1978[7]	12.8°	12.8°	14.2	15.0+	13.6

N.B. Includes HIS sample residing in all areas, not just SMSAs.
* ° = Lowest value on each line.
† + = Highest value on each line.

region. The data in Table 8 omit visits to hospital inpatients.

Substantial interregional differences in health status and mortality have been observed by demographers and epidemiologists for many specific conditions.

More research should be done on the nature of the differences in health status by region of the country. Some differences can be attributed to the life-style in the West. For example, the West is high in the number of accidents. But other differences in morbidity may be due to selective in-migration to the West of those suffering from certain chronic conditions such as arthritis and respiratory problems. In general, it has been assumed that in-migrants tend to be healthier than others in the population. While in-migrants are quite young, they may not be healthy.

Restricted activity days are caused by a wide variety of conditions. Obviously some diagnoses are much more likely to be associated with hospital care than other diagnoses (e.g., a serious injury vs. an upper respiratory infection). Some of this variation in causes of restricted activity days is controlled by the use of age-sex specific categories. For example, in 1975 about 30 per cent of all restricted activity days for those under age six were for upper respiratory infections, while this condition comprised under 10 per cent of the total for the entire population.

In the present analysis it has been tacitly assumed that the mix or morbidity for each given age-sex category is relatively uniform in each region of the country, or at least uniform in its need for hospital care. We have done a study (unpublished) for the 1975 HIS West sample similar to the one reported here. In that study we standardized hospital use by 40 diagnosis-specific morbidity measures for six age-sex groups. The expected hospital use from this diagnostic specific approach gave results for four diverse populations that differed only a few per cent from those based on the method used in this analysis. This supports our assumption regarding morbidity mix. However, our approach probably could be improved by giving different weights to restricted activity days with and without bed disability as well as using other measures of the severity of restricted activity day.

Our use of age-specific restricted activity days as a proxy for the types of conditions which lead to hospitalization also is supported by standards shown in Table 3. Apart from infants under age one, the standards had a much narrower range than age-specific hospital use of persons with and without limitation of activity. This

TABLE 8. Annual Physician Visits Per Person by Age and by Region Residents of SMSAs—1973–1974 Health Interview Surveys*

Age	Region				
	Northeast	North Central	South	West	All U.S.
Under 17	4.6^{+}†	4.3°‡	4.3°	4.3°	4.4
17–44	4.9°	5.1	4.9°	5.8^{+}	5.1
45–64	5.5	5.5	5.4°	6.4^{+}	5.7
65+	7.1	6.6	6.3°	8.0^{+}	7.0
All	5.2	5.2	4.9°	5.7^{+}	5.2

* Adapted from Gentile A. Health characteristics by geographic region, large metropolitan areas, and other places of residence, U.S., 1973–1974. Series 10: Data from National Health Survey, no. 113, (DHEW Publication No. (PHS) 75-1540).

† $^{+}$ = Highest value on each line.

‡ $^{\circ}$ = Lowest value on each line.

suggests that much of the variation in hospital use by different populations can be explained by different restricted activity day rates of the populations.

The days spent recuperating at home after a hospital episode are classed as restricted activity days by the HIS. Thus some restricted activity days are a consequence of hospital care. In our model this raises the expected hospital use in areas where hospital use is high. This effect is most prominent for a hospitalized person who would not have been considered restricted in activity by HIS when he entered the hospital (e.g., for some elective surgery, for some diagnostic admissions).

The magnitude of this effect is small. For example in our 1975 SMSA HIS sample the annual hospital nights were 5 per cent of annual restricted activity days. If the average post-hospital convalescence was equal to the time in the hospital then about 10 per cent of restricted activity days could be attributed to either hospitalization or convalescence from hospitalization. The overall effect of counting hospital convalescence as a restricted activity day would be to reduce and not increase differences in morbidity standardized hospital use by region.

Possible Explanations of Regional Differences in Standardized Ratios

There are many possible explanations of the regional differences in standardized hospital-use ratios. Health care coverage is among them. The West was substantially slower than the midwest and Northeast in enrolling its population in Blue Cross hospital insurance after World War II.[11] The states with high Blue Cross enrollment in 1947 tend to be those with longer average lengths of stay in the 1970s. Even today the West is a little lower than the U.S. average in the percentage of its population under age 65 with private health care coverage. This could lower relative hospital use in the region.

The West is the region with the highest percentage of its population enrolled in prepaid group practices. Our analysis of the 1975 HIS showed that 10.4 per cent of the western SMSA population under age 65 were members of prepaid group practice plans (PGP). Over many years the competition between prepaid group practice and other private coverage in the West may have led to less hospital use in that region. Prepaid group practice plans have two effects on utilization. They have less hospital use by their members (relative to health status) and they also may lower hospital use of those with competing private health care coverage within their service areas. The influence of PGP in the West is a key hypothesis to be tested in seeking an explanation of the observed regional difference of hospital use.

There may be relatively more surgery performed on patients in the West who do not spend a night in the hospital even though they use hospital operating and recovery rooms. Unfortunately, there are not good national statistics on the volume or trends of such nonadmitted surgical cases, but they are quite common in Western prepaid group practices.[12] It also is possible that home health care differs by region of the country, although such care generally is considered a substitute for long-term, not short-term, hospital care.

The West has been a leader in developing insurance for office dental care, which can reduce the demand for inpatient dental surgery. The number of inpatient dental surgery operations per hundred thousand in the Northeast region is 293 while it is only 54 in the West.[13] This is one small example of the influence of health insurance coverage on hospital use.

Some of the regional differences in use, particularly for cases that are not surgically treated, may be attributable to regional differences in providing care of mental illness and other long-term hospital care in short-term hospitals. The Health Interview Survey only determines whether or not a hospital is short term and not whether care was in a long-term unit of a short-term hospital.

At one time, the West had relatively fewer medical schools and teaching hospitals than other regions of the country and even today the West has a smaller proportion of its hospital use in large teaching hospitals than other regions. This may contribute to the shorter length of stay in the West since large teaching hospitals generally have higher lengths of stay than other hospitals, even after adjusting for case mix.

The West has not been short of hospital beds and this is not a likely explanation of low hospital use in the region. For example, in 1978 the average occupancy rates for general hospitals in the U.S. were as follows[14]:

Northeast	80 per cent
North Central	74 per cent
South	72 per cent
West	67 per cent

Many planners have been concerned about the low occupancy in the West since they consider it a lure to overuse. Our data demonstrate that it is clearly possible for a region to have low use, even with substantial empty hospital beds. This is not consistent with the view that bed supply is important in determining use, the so-called Roemer's Law. At best, this law is only half true. Obviously an area cannot use hospital beds that it does not have, but an area need not use its available beds.

Some may believe that regional differences in medical education are the source of interregional variation in hospital use, since physicians admit all patients to hospitals. The majority of physicians practicing in the West, however, received their education at medical schools elsewhere, including a high proportion at schools in north central states. It would be interesting to determine how long it takes a physician with a medical education in the midwest to conform to Western hospitalization conventions when entering practice in the West. In fact, a good deal might be learned on regional hospital use differences by

TABLE 9. Estimated Reduction in U.S. Hospital Nights If All Regions Had the Same Indirectly Standardized Hospital Night Ration As the West Sample Under Age 65 in U.S. SMSAs—1975 Health Interview Survey

		Hospital Nights				
					Annual Use by Sample	
Region	Number of Persons	Annual Expected Rate per 100 Persons	Indirectly Standardized	Annual Actual Rate per 100 Persons	Actual	Expected Indirectly Standardized If Ratio Was .644
	A[*]	B[†]	C[‡]	D[§]	E	F[¶]
Northeast	19,506	68.59	1.094	75.04	14,637	8,616
North central	19,045	63.75	1.276	81.35	15,492	7,819
South	19,158	72.19	0.994	71.76	13,747	8,907
West	14,651	86.76	0.644	55.87	8,186	8,186
All	72,360	71.95	1.000	71.95	52,062	33,528

[*] From line c of Table 4.
[†] From line 1 of Table 4.
[‡] From line j of Table 5.
[§] Column B × Column C.
[‖] (Column D × Column A) Y 100.
[¶] (Column A × Column B × 0.644) Y 100.

surveying practice changes of physicians who go from the North Central to the Western states and those with the opposite migration pattern.

Implications

The wide range of indirectly standardized hospital nights in the four regions of the U.S. raises important questions on the appropriateness of hospital care throughout the country. We cannot determine from our data the amount of hospital care that is appropriate for any person. Therefore, we can only identify populations that are relatively high or relatively low in their observed hospital use adjusted for their reported morbidity.

This study raises cause-and-effect issues on the matter of hospital care. For example, are those residing in the North Central region healthier because they have more hospital care? And, are those in the West less healthy because of less hospital care? It is essential to obtain information on this issue before concluding which region has the most appropriate amount of hospital care. We are inclined to believe that hospital overuse is a greater problem than underuse, particularly for the great majority of the urban population under age 65 that have conventional health coverage. When the patient has more comprehensive coverage for care in the hospital than for care outside of the hospital, both the physician and the patient have incentives to prefer the hospital over the office as a place for care.[15]

The diversity of health care services in the U.S. provides many large scale natural experiments in alternative ways to deliver health care. In order to interpret the influence of different health care delivery systems it is necessary to standardize for the health risks in the various study populations. This paper presents one method of standardizing that may help in the interpretation of the results of these natural experiments.

The method of indirect standardization presented in this paper is applicable to the vast store of data in current and past Health Interview Surveys. The HIS is a potential treasure trove for studying the epidemiology of health care in the U.S. despite some limitations (e.g., no data on decedents or the institutionalized). Modern computer programs simplify making analyses of the type presented here and public-use tapes of the Health Interview Survey are readily available at a modest cost to investigators. Further work could and should be done on the epidemiology of hospital use as it relates to health status, other health care services and socioeconomic variables. For example, the present study should be redone for the nine census divisions of the country. It would be possible to provide reasonably accurate data for a half dozen of the largest metropolitan areas. (Analyses of areas smaller than the four major census regions may require pooling of several years' data to augment sample sizes.) We have already compared the central cities with the balance of SMSAs by region of the country. While the central cities are higher in expected hospital use per capita than their surrounding counties, the ratio of actual to expected hospital use within any given region is very similar for the central cities and their adjacent counties.

One implication of our findings is that the opportunities for modifying hospital use through either regulatory means (e.g., rate regulation), voluntary efforts (e.g., hospital review committees) or market competition vary considerably by region. In essence, a region with a low indirectly standardized hospital-use ratio may have less opportunity to lower its use than another area with a high indirectly standardized ratio. Retrospective evaluations of regulatory and competitive experiments should take into account these regional differences in potential for change.

The PSROs (Professional Standards Review Organization) have sought to determine whether observed differences in hospital length of stay for hospitals within a given area are justifiable, but they have tended to use local standards to make this judgment. We believe that the differences in hospital use between various regions are more important. It might be instructive to have physicians from the west make rounds at north central hospitals (or serve on north central PSROs) and vice versa.

Any national health insurance program should consider regional differences in health status when allocating health care resources. Obviously, less healthy populations warrant more care, all other things being equal. Regional differences in health status should be included as a consideration in the monitoring of government programs such as Medicare and Medicaid. The interregional variations in age specific morbidity in Table 2 demonstrate one problem of using age and sex along as substitutes for health status.

We found more interregional variation in nonsurgical hospital case rates than in surgical case rates. Perhaps some reconsideration should be given to the concept of discretionary hospitalization, which some equate with elective surgery. Elective surgery may be discretionary with respect to the time and place of the operation but some elective surgery is not really discretionary in terms of its necessity. For example, the surgical repair of some severe congenital malformations may be quite necessary even if the surgery is elective in time and place. In contrast, many medical admissions to hospitals may be urgent in time but discretionary in terms of the necessity of hospitalization. Criteria for medical admissions are likely to be less rigorous than those for surgical admissions and hence, more subject to nonmedical considerations by both patient and physician.

The general shift of the U.S. population to the South and West continues. If these in-migrants follow the patterns of hospital care in these two regions, then this population shift should reduce overall hospital care for a U.S. population constant morbidity. This is because people are moving from geographic areas with higher indirectly standardized use ratios to those with lower ratios.

There are a few studies of the secular trends in health status in the U.S. A recent article[16] presents some data from the Health Interview Survey between 1966 and 1976 and concludes that more morbidity is being reported by the National Health Survey Samples over time. Our data in Table 7 lead to the same conclusion. Changes in the age-sex specific morbidity of the population over time may be contributing to health care utilization trends. It is quite possible for disability to be increasing, even while death rates decline. Prolonged life could be associated with more limitation of activity and restricted activity days.

It is tempting to speculate on the cost implications of our finding. If one assumes that the volume of hospital care in the West is at least sufficient (considering the morbidity of its population) then the indirectly standardized hospital ratio of 0.65 in the West implies that the U.S., including the West, would use 35 per cent fewer hospital nights if it conformed to Western practices. Because the logic of this estimate was not obvious to several reviewers, Table 9 was prepared to show its basis. The sum of the regional data in column E is the total actual annual hospital nights used by the sample population. The sum of the regions in Column F is the number of nights that would have been used by the study sample if each region had had an indirectly standardized ratio of 0.644 (i.e. the same as the West). The expected use divided by the observed use is 0.644 (i.e. 33,528 ÷ 52,062). Thus, although there would be no change in the West by adopting Western standards, the average reduction in the other three regions would be greater than 35 per cent.

Since hospital days for the less sick are likely to comprise much of the reduction in use, the percentage that hospital inpatient costs were reduced would be less than the percentage that the use was reduced.[17] Nonetheless, the total dollar savings could be 20 to 25 per cent of all inpatient U.S. short stay hospital costs, currently estimated at about $80 billion per year. In addition, there should be some savings in physician and surgeons costs, as well as a reduction in the iatrogenic morbidity which is a common consequence of hospital care.[18, 19]

Acknowledgments

D. Taya Dunn Mangnall performed all of the computer analyses for this study using the computer resources of the Northern California Region of Kaiser Foundation Health Plan, Inc.

References

1. National Center for Health Statistics. Health United States 1980. Hyattsville, Md.: National Center for Health Statistics, 1980 (DHHS Publication No. (PHS) 81-1232).

2. Wilder CS. Health Characteristics by Geographic Region, Large Metropolitan Areas, and Other Places of Residence, U.S. July 1963–June 1965: National Center for Health Statistics, (Series 10: Data from the National Health Survey, no. 36) (PHS) 1000.

3. Wilder MH. Disability Days, United States–1968, (Series 10: Data from the National Health Survey, no. 67) (DHEW Publication No. (HSM) 72-1034).

4. Namey CW, Wilson RW. Health Characteristics by Geographic Region, Large Metropolitan Areas, and Other Places of Residence U.S. 1969-70, (Series 10: Data from the National Health Survey, no. 86) (DHEW Publication No. (HRS) 74-1513).

5. Wilder MH, Pearson AN. Disability Days U.S., 1971 (Series 10: Data from the National Health Survey, no. 90) (DHEW Publication No. (HRA) 74-1517).

6. National Center for Health Statistics, Health United States 1979, Hyattsville, Md.: National Center for Health Statistics, 1979 (DHEW Publication No. (PHS) 80-1232).

7. National Center for Health Statistics. Health United States 1980. Hyattsville, Md.: National Center for Health Statistics, 1980 (DHHS Publication No. (PHS) 81-1232).

8. Wilder CS. Limitation of Activity Due to Chronic Conditions: U.S., 1974, (Series 10: Data from the National Health Survey, no. 111) (DHEW Publication No. (HRA) 77-1537).

9. Gentile A. Health characteristics by geographic region, large metropolitan areas, and other places of residence, U.S., 1973–1974, (Series 10: Data from National Health Survey, no. 113) (DHEW publication no. (PHS) 75-1540).

10. Wilson RW. Do health indicators indicate health? Am J Public Health 1981;71:461.

11. Reed LS. Blue Cross and medical service plans. Washington, D.C.: Federal Security Agency, 1974.

12. Marks SD, Greenlick MR, Hurtado AV, Johnson JD, Henderson J. Ambulatory surgery in an HMO: a study of costs, quality of care and satisfaction. Med Care 1980;18:127.

13. Ranofsky AL. Utilization of short stay hospitals: annual summary for the United States, 1975. National Center for Health Statistics, (Series 13: Data from the National Health Survey, no. 31) (DHEW publication: No. (HRA) 77-1782).

14. Strahan, GW. Inpatient Health Facilities Statistics U.S., 1979. Data from the National Health Survey Series 14, No. 24. (DHHS publication no. (PHS) 81-1819).

15. Blumberg MS. Health status and health care use by type of private health coverage. Milbank Mem Fund Q 1980;58:633.

16. Colvez A, Blanchet M. Disability trends in the U.S. population 1966–1976: analysis of reported causes. Am J Public Health 1981;71:464.

17. Blumberg MS, Gentry DW. Routine hospital charges and intensity of care: a cross-section analysis of fifty states. Inquiry 1978;15:58.

18. Couch NP, Tilney ND, Rayner AA, Moore FD. The high cost of low frequency events—The anatomy and economics of surgical mishaps. N Engl J Med 1981;304:634.

19. Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at the university hospital. N Engl J Med 1981;304:638.

MEDICAL CARE
September 1982, Vol. XX, No. 9

Surgical Rate Variations: Do They Reflect the Health or Socioeconomic Characteristics of the Population?

NORALOU P. ROOS, PH.D.,* AND LESLIE L. ROOS, JR., PH.D.†

Major unexplained variations in surgical rates have been documented across geographic areas over at least the last four decades. This paper attempts to assess the contribution that population characteristics make to these variations. Variations in surgical rates among the elderly across 56 small rural areas are compared with variations in characteristics of the elderly population of these areas including self-reported health status, levels of disability, mental status, socioeconomic status, ethnic characteristics, etc. The data do not support a needs model for explaining variations in surgical rates. High rate areas were not characterized by an elderly population that was disabled and in ill health. High rate areas did contain a more educated elderly population, as well as more elderly whose national origins were Canadian, British or American.

MAJOR UNEXPLAINED VARIATIONS in surgical rates have been documented across geographic areas over at least the last four decades. Glover[1] found a tenfold variation in British tonsillectomy rates in 1949, ranging from 40 to 440 per ten thousand children. Vayda and his collaborators noted both the existence of these rate variations across Canadian provinces in 1968 and their persistence over the subsequent ten-year period, the period during which national health insurance was introduced.[2,3] Wide intraprovincial variations in surgical rates within Manitoba,[4] Ontario[5] and Saskatchewan[6] also have been demonstrated.

Variations in surgical rates across small areas within the United States have been found to be as large as, if not larger than, those reported in Canada. Wennberg and Gittelsohn[7] reported that one Vermont hospital service area performed tonsillectomies at a rate of 151 cases per 10,000 residents, compared to the neighboring four hospital service areas whose rates ranged between 32 cases and 39 cases per 10,000 residents. Data from Maine[8] indicate widespread variations in surgical rates for nine different surgical procedures (in-

---

* Social and Preventive Medicine and Business Administration, University of Manitoba, Winnipeg, Manitoba, Canada.

† Business Administration and Social and Preventive Medicine, University of Manitoba, Winnipeg, Manitoba, Canada

Interpretations of the data are the authors' own and do not necessarily represent the official opinion of Health and Welfare, Canada, the Manitoba Health Services Commission or the Manitoba Department of Health.

This research was supported by Grant No. 6607-1157-46 from Research Programs Directorate, Health and Welfare, Canada, by a Career Scientist Award No 607-1001-22 to Dr. Noralou Roos and by a Health Scholar Award No. 606-1114-48 to Dr. Leslie Roos.

Reprint requests: Noralou P. Roos, Department of Social and Preventive Medicine, University of Manitoba, 750 Bannatyne Avenue, Winnipeg, Manitoba R3E OW3, Canada.

0025-7079/82/0900/0945/$01.50 © J. B. Lippincott Co.

cluding appendectomy, prostatectomy, inguinal hernia, etc.), while Lewis[9] found three to fourfold variations in Kansas regional rates for the performance of six common surgical procedures.

Finally, surgical rate variations across countries have been documented. Bunker[10] identified the differences in surgical rates between the United States, England and Wales, while Vayda examined the difference[11] and persistence[12] of differences between Canadian surgical rates and those of England and Wales. McPherson et al.[13] have focused on the patterns of regional variations within and between these three countries.

To date, no one questions the existence of these variations. However, the meaning of these variations is not clear. Most explanations focus upon the link between surgical rates and the availability of physician and hospital resources. Studies have shown "a strong positive correlation between an area's physician-population ratio and the per capita utilization of health services over time."[14] As long as there are more surgeons and more hospital beds available to a population, more surgery will be perfomed. Lewis,[9] in an early study of this relationship, concluded that his results "might be interpreted as supporting a medical variation of Parkinson's Law; patient admissions for surgery expand to fill beds, operating suites and surgeons' time." Although Vayda's earlier work[5] supported this law, more recent analyses of the Canadian data show no consistent relationship between the availability of surgical resources and hospital beds, and the amount of surgery performed.[3,12]

Bunker[1] was one of the first to suggest that economic factors may help explain the existence of variations in surgical rates. Incentives operating in a fee-for-service system might contribute to variations in rates observed between prepaid and fee-for-service practices in the United States, and to the differences between Britain and the United States.

However, while financial incentives may contribute to surgical rate variation, later research by Bunker and Brown[15] has confused this issue. Elective surgical rates for physicians and their spouses were found to be as high as or higher than such rates in other professional groups. These results suggest that physicians, as informed consumers of medical care, place a high value on surgical treatment.

Despite the uncertainty about the causes of variations in surgical rates, the simple existence of these variations suggests to some that unnecessary surgery must exist. Goran,[16] a former director of the Office of Professional Standards Review Organizations (PSRO), observed the variations in surgical procedures across PSRO geographic areas and argued that the lowest use represents a target level that all areas should strive to achieve.

Others take strong exception to this position.[17] Pauly[18] thinks that significant variations in the prevalence of conditions that prompt surgery would explain part of the variation in rates. Some have expressed the concern that low British surgical rates may be attributed to a failure to meet the existing health needs of the population.[19] Rutkow and Zuidema[20] ask the important question:

> Might not studies comparing case loads in various countries point out socioeconomic differences concerning the delivery of health care within each country as opposed to a true indicator of unjustified surgical rates?

Unfortunately little research has been done on the relationship between surgical rates, and the health and socioeconomic status of a population. Our tonsillectomy study in Manitoba[21] was one of the few efforts attempting to correlate surgical rates with morbidity rates. No relationship was found between the tonsillectomy rate in an area and the prevalence of respiratory morbidity in the relevant population. Conclusions from this paper were necessarily tentative: only nine regions were examined, morbidity data were obtained from claims rather than from patient self-reports and only one surgical procedure was studied.

Wennberg and Fowler[22] have examined the relationship between surgical rates and population characteristics in a different fashion. They identified six geographically contiguous areas in Vermont known for their variations in surgical rates and conducted three hundred household interviews in each area. Differences in population characteristics across these areas did not explain their different use of health care. Illness rates and the extent of insurance coverage varied little; existing income differences were not related to variation in the consumption of health care. While Wennberg and Fowler concluded that consumers probably do not determine variations in rates of health care, this study also was handicapped by the few areas available (six) as well as by the limited indicators of surgery (only overall surgical rate is discussed) and population morbidity.

The present research examines the relationship between surgical rates and population characteristics over 56 hospital areas in rural Manitoba. This paper focuses on surgery in the elderly (those 65 and over). The elderly are an important group in which to study these questions. Certain types of surgery, including lens removal and hip surgery, offer the ability to function independently. Yet, the elderly patient experiences a higher surgical complication rate, slower recovery and higher postoperative mortality. Therefore, both the positive and negative implications of high and low rates become important.

This paper examines overall operative rates among the elderly, as well as the amount of elective surgery (in general and for one specific procedure) and the rate of complex surgery. Variations in surgical rates across the 56 areas are compared with variations in important characteristics of the elderly—self-reported health status, levels of disability, mental status, socioeconomic status, ethnic characteristics and so forth.

This analysis addresses several important questions in the literature. Pauly[18] has put forward a needs model for explaining the variations in surgical rates. The needs model might predict that the worse the overall health of an area's population, the more need for surgical intervention. The higher the illness rate in an area (as measured from interviews with the elderly in 1971), the more surgery expected to take place in that area. Conversely, the healthier the population, the less surgery expected in that area.

In a more specific test of the relationship between health characteristics and subsequent need for surgery, the needs model also would predict that the more area residents reporting "eye trouble not relieved by glasses," the greater the need for corrective surgery, and hence, the higher the lens-removal rate.

The socioeconomic explanation of surgical rate variations[20] generates different hypotheses about the relationship between consumer characteristics and surgical rates. This model suggests that economic differences are a major contributor to the differences in surgical rates between the United States and Britian. The United States' ability and willingness to spend more money on health care supports more surgery, presumably to the benefit of the population.

Canadians enjoy universal health care insurance, with no deductibles that would affect access to surgery (except for cosmetic plastic surgery). Differences in surgical rates can not be attributed to residents' differential abilities to pay for medical care, although an economist might be concerned about income foregone during convalescence. If individuals with higher incomes and higher education show more sophistication in the use of medical care and place a high value on surgery, areas with disproportionate numbers of highly educated, wealthier elderly should show higher surgical rates. This relationship might be particularly strong for elective surgical rates, characterized by their impact on quality of life, not their life-saving potential.

Method

This research analyzes the relationship between the health and other pertinent characteristics of a small area's elderly population and the area's surgical rate. As developed by Wennberg and his collaborators,[7] small area analysis involves developing population units (to obtain the denominator) and counting all health care utilization by this unit regardless of where it takes place. A denominator counting all individuals living in an area makes it possible to adjust the utilization experience of the population by age and sex, removing a possible patient-related contribution to surgical rate variation.

This paper focuses on surgery among the Manitoba elderly resident outside of Winnipeg, the provincial capital. Winnipeg residents, over half the population of the province, are excluded because service areas in a major metropolitan area are more difficult to define. However, as noted above, any rural Manitoban having surgery performed in Winnipeg will have this surgery attributed back to his home area.

A total of 56 hospital areas have been defined. Five of the areas contain two hospitals, while 51 have one hospital. Table 1 summarizes the most important characteristics of the 56 hospital areas. Because of the range in most of the numbers making up the table, reporting the median (the case lying exactly in the middle once all the cases have been rank ordered from highest to lowest), rather than the mean, seemed more appropriate.

The rural hospitals are mainly small (the median size being 24.5 beds) and have low occupancy rates (the median rate being 76 per cent). Rural Manitoba also has an ample supply of hospital beds, the median rate per thousand population falling at 6.3. row D of Table 1 reports the percentage of patients in municipalities assigned to a specific hospital or hospitals that receive their care within the hospital(s) in their area. The area usage figures are somewhat lower than those usually reported by Wennberg, partly because of rural patients' heavy usage of Winnipeg hospitals. (A high percentage of Manitoba's specialists are located in Winnipeg.) The figures in rows D and E were calculated using the entire population of the hospital areas, not just the elderly population.

Row F records the size of the target population in the 56 areas. While the num-

TABLE 1. Characteristics of 56 Hospital Areas

	Median	Minimum	Maximum
Hospital Characteristics			
A. Hospital size (# of Beds)	25.5	7	219
B. Occupancy rates	76.2%	49.3%	93.4%
C. Beds/1000 population	6.3	3.1	23.7
D. Per cent of patients in area who use their area hospital (vs. going elsewhere)	57.3	32.6	85
E. Per cent of patients using area hospitals who are from area	70.9	26.7	100
Size of Target Population in Area			
F. Size of elderly population (66+) in each area	541	141	3713
G. Number of individuals in area interviewed in 1971	37	6	395

bers are not large, all analyses have been run with and without the 11 areas having fewer than 250 elderly. Rates in these areas might be expected to be least stable. Results are reported, where noticeably different.

Results

Interviews With the Manitoba Elderly

In 1971 extensive interviews with a large probability sample of Manitoba's elderly were conducted.[23,24] This survey included residents in the community, hostels, personal care homes, extended care units and mental institutions. Only the very few elderly who were long-term residents of acute hospitals were excluded from the sampling frame.

A 5 per cent probability sample was drawn from persons 65 and over living in rural communities (not in a facility). Those living in facilities (nursing homes, hostels and public housing units) were sampled at a rate of 20 per cent. In calculating the characteristics of the areas from interview data, the numbers have been reweighted to approximate the general population.

Two parallel samples were drawn using the master registry of the Manitoba Health Services Commission. When an individual could not be located or refused to participate, a counterpart from the second sample was randomly assigned to help maintain appropriate sampling for statistical purposes. Seventy-two per cent of those interviewed were drawn from the first sample; 28 per cent of the respondents came from the second sample. No significant differences were found between the two samples along several important demographic (marital status, etc.) and utilization (hospital admissions, etc.) characteristics.[25]

The sample appears to accurately reflect several characteristics (age, sex and marital status) of Manitoba's elderly population. The sample somewhat underrepresents the more educated elderly (33 per cent of the sample, as compared to forty per cent of the province's elderly, report nine or more years of education) and underestimates the rate of hospitalization by approximately 15 per cent. This bias, even in a study using proxy interviews, multiple callbacks and an all-inclusive sampling design, suggests the difficulty of accurately identifying a representative elderly sample. A total of 2,626 elderly were interviewed from our 56 rural hospital areas.

Health Status

The interview itself contained 517 separate measures of personal, social, health and demographic characteristics. This analysis focuses on a small subset of the best-known and validated measures. Six survey measures of health status are presented in the top part of Table 2. For each measure, the mean and standard deviation are reported across the 56 areas. When the smallest areas are excluded, the mean never changes by more than 2 percentage points, although the standard deviation (as one would expect) is reduced by up to 25 per cent. Most of the items are self-explanatory; both self-perceived health status and the number of health problems have been shown to predict health services utilization (number of physician visits and hospitalizations).[25]

The next to last health status measure is the Shanas Disability Index which predated the Katz[26] activities of daily living index; the two contain similar types of items. Both the Katz and the Shanas indices have been criticized for not measuring whether an individual needs help with various activities, but whether the individual is receiving help (needed or not). This may be a particular problem where an elderly individual is living with a spouse. We have attempted to eliminate this bias by recalculating the Disability Index, including only those area respondents living alone. The last health status measure, therefore, reports the per cent of those elderly interviewed in an area who live alone and show

TABLE 2. Additional Characteristics of the Areas' Elderly Population: Health Status, Mental Status, Life Satisfaction, Access to Health Care and Socioeconomic Status

	Mean	Standard Deviation
Health Status		
Per cent of respondents in area reporting that health is good or excellent for their age	59.5	13.71
Per cent of respondents in area reporting none of a series of 15 health problems they were asked about, including such things as stomach trouble, arthritis etc.	12.8	7.69
Per cent of respondents reporting no eye problems	79.4	9.78
Per cent of respondents in area reporting spending no time in bed sick in the last year	83.1	9.85
Per cent of respondents in area showing no functional disability from responses to 13 questions including do you need help with dressing, medication, housework, managing stairs, etc.	80.9	7.97
Per cent of respondents living alone who show no functional disability	84.2	14.94
Life Satisfaction		
Per cent of respondents in area scoring Excellent (17 or higher) on life satisfaction index (20 agree-disagree questions including "I feel old and somewhat tired")	26.6	14.22
Access to Health Care		
Per cent of respondents in area saying no difficulty seeing doctor (or dentist)	89.8	10.88
Per cent of respondents in area who report having regular physician	89.1	8.20
Socioeconomic Status		
Per cent respondents in area with 9 or more years education	27.2	15.41
Per cent of respondents in area who say income and assets currently satisfy their needs very well or adequately	76.0	12.01
Per cent of respondents in area whose income was above official poverty line*	16.3	8.5
Ethnic Status		
Per cent of respondents in area whose national origins were Canadian, American or British	52.8	28.50

* This variable was based only on the responses of those elderly living in the community or in public housing units.

no functional disability. The mean for this measure is somewhat higher than that for the previous measure, suggesting either that disability indices overestimate the disability in individuals not living alone, or that individuals may be able to live alone simply because they are less disabled.

The measures of health status are intercorrelated. Areas with a high percentage of respondents reporting good health were also areas where respondents were more likely to report no health problems (r = 0.44, sig 0.001).[‡] Such areas also were more likely to have respondents report to have spent no time being sick in bed in the last year (r = 0.28, sig 0.02). However, the self-reported health status of respondents in an area and the disability indexes were not related.[§]

Life Satisfaction, Access and Socioeconomic Status

The index of life satisfaction is a thoroughly researched, widely used measure in the elderly population.[28,29] To the degree that elective surgery might be sought by individuals wanting to improve the quality of life, low satisfaction scores might be associated with high surgical rates.

Two measures of access to health care include the percentage of respondents in an area reporting no difficulty in seeing a doctor (or a dentist) and the percentage of respondents in an area having a regular physician. These measures were correlated across the areas (r = 0.33, sig 0.01). The association (or lack thereof) of these variables with surgical rates will help test the Rutkow-Zuidema[20] suggestion that low surgical rates occur where the population has restricted access to medical care.

The three measures of socioeconomic status include the percentage of respondents in an area with nine or more years of education, a subjective measure of whether the respondents' income satisfies their needs, and an objective measure of the percentage of elderly in an area with incomes above the official poverty line. In calculating this latter measure we have excluded individuals who were in nursing homes. Such individuals have their food and shelter provided and have quite different income needs than those living in the community.

As Table 2 indicates, although 76 per cent of respondents across areas reply that their income and assets currently satisfy their needs very well or adequately, only 16 per cent (on average across areas) of the respondents were above the poverty line in 1971. (The 1971 Canadian poverty line figures were $204 a month for single adults, and $375 a month for a family of two.) While these figures reflect the elderly's dismal economic plight, some of the rural elderly live on farms and have income in kind not reflected by these figures.[‖]

The final measure in Table 2 indicates the ethnic heterogeneity of the area. The standard deviation of this measure (the per cent of respondents with Canadian, American or British national origins) reflects Manitoba's ethnically distinct communities. It will be possible to determine if Anglo-Saxons use surgical care differently from other ethnic groups.[¶]

Surgical Rates

Data on surgical rates are taken from the hospital discharge files of the Manitoba

---

[‡] Note, the Pearson correlation coefficient used throughout this paper is based on an N of 56 hospital areas unless otherwise noted. Since we are dealing with population-based surgical rates, not sample data, the level of significance has no universally accepted meaning.[27]

[§] In a separate analysis, an area's overall hospitalization rate as judged from the claims data was strongly associated with the amount of hospitalization reported by those being interviewed in an area. This was a useful check on the validity of the two measures in the different data sources.

[‖] As a partial check, small area census data on education (mean years of education) and income (mean individual and mean household income) for the whole adult population were compared with the survey data from the elderly sample. The correlation between the two data sets was 0.68 for the education variables. The census income variables showed reasonable correlations with the objective measure of poverty among the elderly (0.47 and 0.44), but not with the subjective income measure (0.18 and 0.24). The subjective and objective income measures of income also correlated poorly at the individual level.

[¶] Adult census data also were used as a partial check on this survey measure. The census data on percentage of English-speaking people in each area were highly correlated (0.76) with the survey measure of Canadian, American or British national origins.

Health Services Commission. The whole population of Manitoba is covered by provincially administered health insurance. Hospital and medical care are documented in considerable detail because of the fee-for-service payment arrangement for medical services.

The data base has been fully described elsewhere.[30,31] The quality of data taken from health insurance claims has been the focus of intensive investigation. The data appear to be an accurate, reliable and valid representation of events in the health care system. In tonsillectomy research, claims diagnoses were logically consistent with surgical procedures subsequently or previously performed. More recently, the claims filed by surgeons for operative procedures have been compared with the hospital discharge claims filed independently by hospitals. A high level of correspondence between information contained in the two files was found. Over 90 per cent of the claims filed by surgeons for hysterectomies were confirmed (by date and type of procedure) in the hospital file, and vice versa.

While the interviews are based upon a sample of the elderly interviewed in 1971, the rate data encompass the entire surgical experience of the areas' elderly population. To generate more stable surgical rates by increasing the size of the numerator, two years of surgery (1971–1972) were combined; an average surgical rate/1000 elderly then was calculated. Moving from rates based on one year's data to those produced from two years' data improves statistical stability; the standard deviation is reduced from 23.6 to 16.6 and correlations with subsequent periods were increased. The rates for 1971–1972 were correlated 0.38 with 1974–1975 rates using the Pearson correlation coefficient (significant at 0.002); excluding the 11 smallest areas with 250 or fewer elderly, the coefficient is 0.54 ($p = 0.0001$). Four years' data (1970–1973) on cataract surgery were used. The cataract surgery rates for 1970–1973 were correlated 0.45 with 1974–1977 rates ($p = 0.0001$); with the smallest areas excluded, the correlation was 0.43 ($p = 0.01$). While larger areas would produce more statistically stable rates, making more than one year's data unnecessary, the smaller areas used here are more likely to capture important practice pattern and population differences.

A surgical discharge was defined as any discharge that included one or more "treatment" surgical procedures (a numeric three-digit International Classification of Diseases Adapted—8 surgical code). The rates calculated for each area included the number of individuals coming to surgery, the number of discharges containing one or more surgical procedures and the hernia equivalent (H.E) surgical rate, a weighted measure based upon the California Index of Relative Values developed by E.F.X. Hughes.[32] A fourth measure, the number of discharges with surgical procedures having a hernia equivalent of 0.5 or greater, is used most frequently in this report. This cutting point excludes such minor surgery as D & C procedures (H. E. weight of 0.43), myringotomies (H.E. 0.09), etc., but includes more complex procedures such as appendectomy (H.E. 1.14), inguinal hernia repair (H.E. 1.0) and so forth. The four variables were intercorrelated 0.75 (sig 0.001) or higher. All surgical treatment, whether performed on an inpatient or outpatient basis, was available for this analysis. However, since on average outpatient surgery only contributed 6 per cent to an area's total complexity score, analysis has been restricted to inpatient surgery.

Table 3 reports the range of variation in the measures of surgical treatment. All rates have been adjusted by age and sex, using the indirect method. The first four measures summarize the amount of surgery in each of the 56 areas; rates varied 2.4- to 2.8-fold across the areas.

The amount of elective surgery also differed from area to area: from 17.6 to 56.5

TABLE 3. Surgical Utilization Characteristics of Hospital Areas

	Median	Minimum	Maximum
All Surgical Treatment			
Number of individuals coming to surgery	85.1	53.3 (60.3)*	144.2
Number of discharges with one or more surgical procedures	96.5	67.8 (69.7)	164.0
The hernia equivalent surgical rate (total complexity)	133.2	82.1 (98.7)	225.9
Number discharges with surgical procedures having H.E. GE 0.5	75.7	52.0	124.0
Elective Surgical Rate †	32.6	17.6 (20.6)	56.5
Rate of Performing Complex Surgery (H.E. of 2.0 or greater)	32.2	14.3 (20.4)	53.5
Cataract Surgery Rate ‡	7.2	3.1	13.1

* When excluding the areas with 250 or fewer elderly changes the minimum or maximum, the new figures are listed in parentheses.

† Elective surgery includes the following procedures: Lens extractions (14.1–14.5); prostatectomy (58.1–58.3); hysterectomy (69.1–69.4, 70.1–70.2); cholecystectomy (43.5); hemorrhoidectomy (51.3); varicose vein stripping (24.4); inguinal hernia repair (38.2). All numbers in parentheses refer to ICDA-8.

‡ Annual rate is based on average of 4 years' observations, 1970–1973.

per 1,000 elderly. Seven of the procedures most commonly referred to as discretionary or elective are included in this variable. By restricting the analysis to the procedures having a hernia equivalent of 2.0 or greater, the complex surgery measure counts the following kinds of major surgery; vagotomy and pyloroplasty (H.E. = 2.00), bilateral nephrectomy (4.29), bilateral orchiopexy (2.14), bilateral fixation of testis (2.4), etc.

The final measure records the variation in cataract surgery across the 56 areas. Over the four-year period examined, residents of high rate areas were on the average almost four times as likely to receive this procedure as were residents of low rate areas.

Table 3 showed the range of variation in, and the rate of, the elderly's exposure to different types of surgery. Table 4 outlines the wide scope of surgical practice patterns from a different perspective. Here the areas are characterized according to the tendency to experience one type of surgery versus another. While in Table 3 the elderly population was the denominator for the rate calculations, in Table 4 the number of 1971–1972 surgical procedures

TABLE 4. Patterns of Surgical Practice Percentage of All Surgery (H.E. 0.5 or greater) in Area That Is of the Specified Type

	Median	Minimum	Maximum
Per cent of all surgery that is:			
Elective*	40.3	28	68 (62)†
Complex (H.E. of 2.0 or greater)	40.3	27 (32)	58 (55)

* All variables are defined as in Table 3.

† When excluding the areas with 250 or fewer elderly changes the minimum or maximum, the new figures are in parentheses.

with a hernia equivalent 0.5 or greater is the denominator. The first row suggests that in at least one area, 68 per cent of all surgery (with a hernia equivalent score of 0.5 or more) consisted of one of the seven elective procedures; in another area, only 28 per cent of the surgery was elective.

Surgical Rates and Population Characteristics

Pearson product moment correlations between the characteristics of the elderly interviewed in an area in 1971 and the area's 1971–1972 surgical rates were examined. The data do not support a needs model for explaining variations in surgical rates. High rate areas were not characterized by an elderly population that was disabled and in ill health. In fact, no consistent stable relationships between the health status variables and the surgical rate variables were found. Those statistically significant relationships that did appear (between surgery and no time sick in bed) vanished when the smallest areas were eliminated. The number of elderly reporting eye problems in an area was also unrelated to the frequency of cataract surgery. Grouping the areas according to 1971–1972 surgical rates, as is done in Table 5, further indicates weak and inconsistent relationships.

Correlations between our best quality of life indicator, the widely used life satisfaction index, and the surgical measures, including the elective surgical rate, generally were not statistically significant. This lack of relationship is important, because the major benefit derived from several elective procedures is thought to be quality of life; benefits in terms of years of life saved or cost reduction are difficult to document.[33]

Subjectively perceived access also was unrelated to surgical rates. Some remote rural areas (in 1974 there was 1,105 population per physician in rural Manitoba[34]) may experience a shortage of physicians, but the substantial amount of surgery done in Winnipeg on rural residents makes it difficult to conceive of access in terms of local physician supply alone. Since very few elderly reported problems seeing a physician and most had a regular doctor, barriers to access are unlikely explanations for low surgical rates.

Although neither of the income variables was significantly associated with the surgical rate measures, Table 6 shows that education (the percentage of elderly respondents with nine years or more education) and ethnicity (the percentage of respondents with Canadian, American or British national origins) were strongly and consistently associated with the measures of overall surgical rate.# In those areas with the highest surgical rates, 38 per cent of the respondents claimed nine or more years of education; in the other areas, 26 per cent or less were so educated (Table 5). The ethnic differences vary even more sharply with overall surgical rates. In the lowest rate areas only 41 per cent of the respondents noted their national origins as Canadian, British or American; in the high rate areas, 67 per cent reported such backgrounds. However, neither ethnicity nor education was associated with the two more specific surgical measures: rate of elective surgery and rate of cataract surgery.

Discussion

Either the health characteristics of an area's population are not related to the type and amount of surgery performed on residents, or the measures used for this study are too insensitive to pick up this relationship. Since our indices have been accepted widely for measuring health status and quality of life among the elderly, the latter explanation is unlikely. Even the quite specific data on problems with vision

---

# Generally similar, but slightly weaker, correlations were found using the census data on education, income and ethnicity.

TABLE 5. Illustrating the Relationship Between 1971 Characteristics of an Area Population and 1971–1972 Surgical Rates

	Areas Grouped According to 1971–1972 Surgical Rates*			
Health Status	Areas With Lowest Rates	Medium Low	Higher	Areas With Highest Rates
Health Status				
Mean per cent of respondents in areas reporting their health is good or excellent for their age	59	55	62	65
Mean per cent of respondents in areas reporting none of a series of 15 health problems they were asked about including such things as stomach trouble, arthritis, etc.	12	12	14	14
Mean per cent of respondents in areas reporting spending no time in bed sick in the last year	81	81	85	88
Mean per cent of respondents in areas showing no functional disability from responses to 13 questions including do you need help with dressing, medication, housework, managing stairs, etc.	81	80	84	79
Mean per cent of respondents living alone who show no functional disability	83	84	83	88
Life Satisfaction				
Mean per cent of respoadents in areas scoring excellent (17 or higher) on life satisfaction index (20 agree-disagree questions including "I feel old and somewhat tired")	25	24	29	30
Access to Health Care				
Mean per cent of respondents in areas saying no difficulty seeing doctor (or dentist)	88	89	92	93
Mean per cent of respondents in areas who report having regular physician	88	87	90	93
Socioeconomic Status				
Mean per cent of respondents in areas with 9 years or more of education	24	26	26	38
Mean per cent of respondents in areas who say income and assets currently satisfy their needs very well or adequately	75	74	78	77
Mean per cent of respondents in areas whose income was above official poverty line	17	18	15	16
Ethnic Status				
Mean per cent of respondents in areas whose national origins were Canadian, American or British	41	47	63	67
Number of Elderly Interviewed	671	594	576	785
Number of Elderly in the Population	10538	10048	9372	9328

* Areas grouped according to their age and sex adjusted rate of the number of surgical discharges with hernia equivalent score greater or equal to 0.5.

TABLE 6. Significant Correlations Between Characteristics of an Area's Elderly Population and the Area's Surgical Rates (Age and Sex Adjusted per 1000 Elderly)

	Individuals Having Surgery	All Surgical Discharges	H.E. Surgical Rate	Discharges With H.E. 0.5 or More	Elective Surgery	Complex Surgery	Cataract Surgery
Socioeconomic Status							
Per cent of respondents in area with 9 or more years of education	0.45§	0.43§	0.34†	0.32§		0.28*	
Per cent of respondents in area who say income and assets currently satisfy their needs very well or adequately							−0.25*
Per cent of respondents in area whose income was above official poverty line							
Ethnicity							
Per cent of respondents in area whose national origins were Canadian, American, or British	0.43§	0.40‡	0.26*	0.23*		0.24*	

Levels of Significance of Pearson R:
* = .05.
† = .01.
‡ = .001.
§ = .0001.

showed no systematic relationship to fourfold variations in cataract surgery rates. The data here accord with several observations by Fuchs.[35] In industrialized societies, the relationship between medical inputs, such as surgery, and presumed outputs, such as health, is tenuous. Although we have not measured health manpower directly, the findings seem in line with Newhouse and Friedlander's[36] results showing no systematic link between the amount of health resources and a number of medical indicators of ill health. Given these data, Pauly's[18] supposition that "there may be significant variations in the (unknown) prevalence of conditions that prompt surgery which can explain part of the wide variation in surgery rates" seems unlikely. Although Barker[37] has clearly documented geographic variations in disease in Britain, he also suggests that such variations are more likely to be reflected in autopsy than in surgical rate statistics.

The relationships among income, educational status and surgical rates are less in accord with the literature. The American study of Bombardier et al.[38] previously found income to be positively associated with surgery, although, other things being equal, the more highly educated tended to have fewer operations. Our differences may be due to the nature of the health care system. Income seems a less important variable in Canada, given the insured care available under National Health Insurance. Bombardier et al. also reported the negative correlation between education and surgery to be least apparent in the oldest and the youngest in their sample. Finally, since in our sample education and ethnicity are highly intercorrelated, the relationship between education and surgery may be spurious.

The correlation between the ethnic makeup of a small area's population and surgical rates is an interesting one. Our data suggest that areas with high surgical rates are also those likely to have disproportionately large numbers of elderly of Anglo-Saxon (English-Canadian or American) parentage. Whether such individuals have a cultural bias towards surgery or are more sophisticated consumers of surgical services (the Bunker-Brown model) can not be determined from these data. In fact, a quite different causal mechanism may be operating. Perhaps North American trained physicians have a surgical bias and are more likely to settle in rural areas having a congenial ethnic makeup. Our ongoing study will investigate the relationship among supply variables, surgical rates and mortality rates. In any case, as Bunker and Brown[15] have observed, supply variables and population characteristics would seem more likely to influence "demand for surgery" than be related to "need for surgery."

The importance of being able to distinguish between demand for surgery (whether influenced by patients or physicians) and the "need" for surgery is clear. High surgical rates have clear short-term cost implications. Wennberg[39] has argued that a high-surgery-use strategy (based on costs in high rate areas) adopted across the U.S. population would cost $6.8 billion per year, although a low-use strategy would cost only $3.0 billion (in 1973 dollars). Since the high-use strategy also may have implications for mortality rates,[40, 41] the importance of establishing the efficiacy of this strategy (or lack thereof) in terms of quality of life is self-evident.

Acknowledgments

The author would like to thank the Manitoba Health Services Commission for facilitating access to claims data and to thank Betty Havens of the Department of Health and Social Development for providing access to the Manitoba Needs Survey. John Bunker, Jack Wennberg and Benjamin Barnes provided valuable suggestions concerning this research.

References

1. Glover JA. Tonsillectomy in the school medical service: increased incidence in 1948. Monthly Bulletin Ministry of Health, London 9:62.

2. Vayda E, Anderson GD. A comparison of provincial surgical rates, 1968. Can J Surg 1975;16:18.

3. Mindell WR, Vayda E, Cardillo B. A ten year study of surgery in Canada 1968–1977. Can Med Assoc J 1982. (in press)

4. White Paper on Health Policy, Winnipeg, Manitoba Department of Health and Social Development: Government of Manitoba, 1972.

5. Stockwell H, Vayda E. Variations in surgery in Ontario. Med Care 1979;17:390.

6. Dyck FJ, Murphy FA, Murphy JK, et al. Effect of surveillance on the number of hysterectomies in the province of Saskatchewan. N Engl J Med 1977; 296:1326.

7. Wennberg JE, Gittelsohn AM. Small area variations in health care delivery. Science 1973;182:14.

8. Wennberg JE, Gittelsohn AM. Health care delivery in Maine I: patterns of use of common surgical procedures. J Maine Med Assoc 1975;66:5.

9. Lewis CE. Variations in the incidence of surgery. N Engl J Med 1969;281:880.

10. Bunker JP. Surgical manpower: a comparison of operations and surgeons in the United States and in England and Wales. N Engl J Med 1970;283:135.

11. Vayda E. Comparison of surgical rates in Canada and in England and Wales. N Engl J Med 1973;289:1224.

12. Vayda E, Mindell WR, Rutkow IM. A decade of surgery in Canada, England and Wales and the United States. Arch Surg 1982;117:846.

13. McPherson K, Strong PM, Epstein A, Jones L. Regional variations in the use of common surgical procedures; within and between England and Wales, Canada and the United States of America. Soc Sci Med 1981;273:15A.

14. Reinhardt UE. The physician as generator of health care costs. Health care and the American economy: issues and forecasts, 1978. Chicago, Health Services Foundation, 1979.

15. Bunker JP, Brown BW, Jr. The physician-patient as an informed consumer of surgical services. N Engl J Med 174;290:19.

16. Goran M. The evolution of the PSRO hospital review system. Med Care 1979;(17 supplement:5).

17. Moore FD. Small area variations in health care delivery; a critique. J Maine Med Assoc 1977;68:49.

18. Pauly MV. What is unnecessary surgery? Milbank Mem Fund Q 1979;57:1.

19. Pearson RJC, Smedby B, Berfenstam R, Logan RFL, Burgess AM, Peterson OL. Hospital caseloads in Liverpool, New England and Uppsala Lancet 1968;2:559.

20. Rutkow IM, Zuidema GD. Unnecessary surgery: an update. Surgery 1978;84:671.

21. Roos NP, Roos LL Jr, Henteleff PD. Elective surgical rates—do high rates mean lower standards? N Engl J Med 1977;297:360.

22. Wennberg JE, Fowler FJ, Jr. A test of consumer contribution to small area variations in health care delivery. J Maine Med Assoc 1977;68:275.

23. Mossey JM, Havens B, Roos NP, Shapiro E. The Manitoba longitudinal study on aging. Description and methods. Gerontologist 1981:21:551.

24. Manitoba Department of Health and Social Development. Aging in Manitoba, vol 1–9. Winnipeg, Manitoba Department of Health and Social Development 1973–1977.

25. Roos NP, Shapiro E. The Manitoba longitudinal study on aging: preliminary findings on health care utilization by the elderly. Med Care 1981;19:644.

26. Katz S, Ford AB, Downs TD, Adams M, Rusby DI. Effects of continued care: a study of chronic illness in the home. U.S. Department of Health, Education, and Welfare, December 1972. (DHEW publication. (HSM)73-3010.)

27. Winch RF, Campbell DT. Proof? No! Evidence? Yes! The significance of tests of significance. Am Sociol Rev 1969;4:140.

28. Berkman L. Measurement of mental health in a general population survey. Am J Epidemiol 1971;2:105.

29. Adams D. Correlates of satisfaction among the elderly. Gerontologist 1971;11:64.

30. Roos NP, Henteleff PD, Roos LL, Jr. A new audit procedure applied to an old question: is the frequency fo T and A justified? Med Care 1977;15:1.

31. Roos LL, Jr, Roos NP, Nicol P, Johnson C. Using administrative data banks for research and evaluation. Eval Q 1979;3:236.

32. Hughes EFX, Fuchs VR, Jacoby JE, Lewit EM. Surgical work loads in a community practice. Surgery 1972;71:315.

33. Bunker JP, Barnes BA, Mosteller F. Costs, risks and benefits of surgery. Oxford University Press, New York 1977.

34. Roos NP, Gaumont M, Horne JM. The impact of the physician surplus on the distribution of physicians across Canada. Canadian Public Policy 1976;2:169.

35. Fuchs VR. Economics, health, and post-industrial society. Milbank Mem Fund Q 1979; 59:153.

36. Newhouse JP, Friedlander MJ. The relationship between medical resources and measures of health: some additional evidence. Rand document R-2066-HEW. Santa Monica, Calif. Rand, 1977.

37. Barker DJP. Geographical variations in disease in Britain. Br Med J 1981;283:398.

38. Bombardier C, Fuchs VR, Lillard LA, Warner KE. Socioeconomic factors affecting the utilization of surgical operations. N Engl J Med 1977;297:699.

39. Wennberg JE. Factors governing utilization of hospital services. Hos Prac 1979;14.

40. Wennberg JE, Bunker JP, Barne B. The need for assessing the outcome of common medical practice. Annu Rev Public Health 1980;1:277.

41. Roos NP, Roos LL, Jr. High and low surgical rates: risk factors for area residents. Am J Public Health 1981;71:591.

C5170-1

Medical Care

October 1982
Vol. 20 No. 10
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

Quality, Cost, and Health: An Integrative Model

Physician Practice Profiles for CME: Four Investigations Into Sampling Methods

Improving Compliance Among Hypertensives: A Triage Criterion With Cost-Benefit Implications

Social Networks, Beliefs, and the Decision to Seek Medical Care: An Analysis of Congruent and Incongruent Patterns

Assessing the Effectiveness of Neonatal Intensive Care

Improving Physician Compliance With Preventive Medicine Guidelines

A Description of a Means of Improving Ambulatory Care in a Large Municipal Teaching Hospital: A New Role for Nurse Practitioners

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Marion Athearn
Jennifer Gardon
Editorial Assistants

S. E. Berki, M.A.
University of Michigan

Robert H. Brook, M.D., Sc.D.
University of California
Los Angeles

Gordon H. De Friese, Ph.D.
University of North Carolina

Eugene Feingold, Ph.D.
University of Michigan

Brant E. Fries, Ph.D.
Yale University

Thomas R. Godkins, B.S., P.A.
University of Oklahoma

Harold Hunter, Dr. P.H.
Washington, D.C.

Sherman A. James, Ph.D.
University of North Carolina

Anthony L. Komaroff, M.D.
Brigham and Women's
Hospital, Boston

James P. LoGerfo, M.D., M.P.H.
University of Washington

Nancy Milio, Ph.D.
University of North Carolina

Duncan Neuhauser, Ph.D.
Case Western Reserve University

John E. Ware, Jr., Ph.D.
The Rand Corporation

Alonzo S. Yerby, M.D.
Harvard University

Alfred Yankauer, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year for individuals, $48.00, for institutions, $55.00; special price U.S. students, $35.00; all other countries except Japan, one year, for individuals, $59.00, for institutions, $68.00. Single copies $6.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, MD, and additional mailing offices.

October 1982
Volume 20, Number 10

CONTENTS

ORIGINAL ARTICLES

Quality, Cost, and Health: An Integrative Model. *Avedis Donabedian, John R. C. Wheeler and Leon Wyszewianski* 975

Physician Practice Profiles for CME: Four Investigations Into Sampling Methods. *Lynn Curry and Wayne Putnam* 993

Improving Compliance Among Hypertensives: A Triage Criterion With Cost-Benefit Implications. *Steven R. Eastaugh and Myron E. Hatcher* 1001

Social Networks, Beliefs, and the Decision to Seek Medical Care: An Analysis of Congruent and Incongruent Patterns. *Emil Berkanovic and Carol Telesky* 1018

Assessing the Effectiveness of Neonatal Intensive Care. *Peter P. Budetti and Peggy McManus* 1027

Improving Physician Compliance With Preventive Medicine Guidelines. *David I. Cohen, Benjamin Littenberg, Cheryl Wetzel and Duncan vB. Neuhauser* 1040

A Description of a Means of Improving Ambulatory Care in a Large Municipal Teaching Hospital: A New Role for Nurse Practitioners. *Daniel M. Becker, Arthur M. Fournier and Laurence B. Gardner* 1046

COMMUNICATION

Spending Time With Patients: The Impact of Organizational Structure on Medical Practice. *Fredric D. Wolinsky and William D. Marder* 1051

BOOK REVIEWS

Learning to Live With Angina, by Cynthia B. Wong. *Thomas S. Inui, Benjamin A. Lipsky and Walter J. Kupersmith* 1060

Response to Inui, Lipsky and Kupersmith. *Cynthia B. Wong and Judith C. Watkins* 1061

Response to Wong and Watkins. *Thomas S. Inui* 1062

The Incidence and Economic Costs of Major Health Impairments: A Comparative Analysis of Cancer, Motor Vehicle Injuries, Coronary Heart Disease, and Stroke, by Nelson Hartunian, Charles N Smart and Mark S. Thompson. *David C. Warner* 1062

Medical Economics and Health Finance, by Steven R. Eastaugh. *S. E. Berki* 1064

Rural Medicine: Obstacles and Solutions for Self-Sufficiency, by Stanley S. Wallack and Sandra E. Kretz. *Roger A. Rosenblatt* 1064

PUBLICATIONS RECEIVED 1067

MEDICAL CARE
October 1982, Vol. XX, No. 10

Improving Compliance Among Hypertensives: A Triage Criterion With Cost-Benefit Implications

STEVEN R. EASTAUGH, SC.D.,* AND MYRON E. HATCHER, SC.D.†

Triage method based upon individual patient characteristics is developed to determine the most effective health education interventions for hypertensive patients. The sample included 402 patients randomly assigned to experimental and control groups. The efficacy of a triage process, whereby patients are subdivided into groups that are more predisposed to benefit from a given health education approach, is discussed. From a limited cost-benefit evaluation, it appears that the benefits of the triage method for achieving medication compliance clearly outweigh its costs only in the case of the highly depressed patients (24 per cent of the sample). Different levels of depression were found to be differentially responsive to different health education treatments for achieving adherence to medication. The most depressed patients were more responsive to family member reinforcement, a 156 per cent increase in the number of high compliers as compared to no intervention.

MUCH INTEREST EXISTS today in targeting public health programs to yield the maximum return on investment. The general objective of this study is to develop patient-specific decision rules that serve as a basis for selecting the health education program most likely to encourage adherence to antihypertensive medication. Decision rules are formed by empirically studying the relationships between salient patient characteristics and treatment effectiveness. A specific objective of the design is to measure the potential importance of mental health variables, such as depression, in explaining the effectiveness of health education intervention efforts. The basic research question is whether interaction exists between the effectiveness of different intervention programs and the characteristics of patients participating in the interventions. If there is a significant link between characteristics and effectiveness, then maximum benefits might be achieved if patients are triaged into the high-yield intervention strategies.

There are several reasons for the current impetus to study efficiency in health education programs. The first reason is the

---

* Assistant Professor of Health Economics, Sloan Program of Hospital and Health Services Administration, Graduate School of Business and Public Administration, Cornell University, Ithaca, New York.

† Director and Senior Associate of Health Systems and Resource Planning Associates, Inc., 16026 Burton Lane, Laurel, Maryland.

Supported by National Institutes of Health Grants HL-17016, MH-14567, and 1-32(HL)-07180.

Presented in part to the Operations Research Society of America and the Institute of Management Sciences Joint National Meeting in Washington, D.C., May 7, 1980.

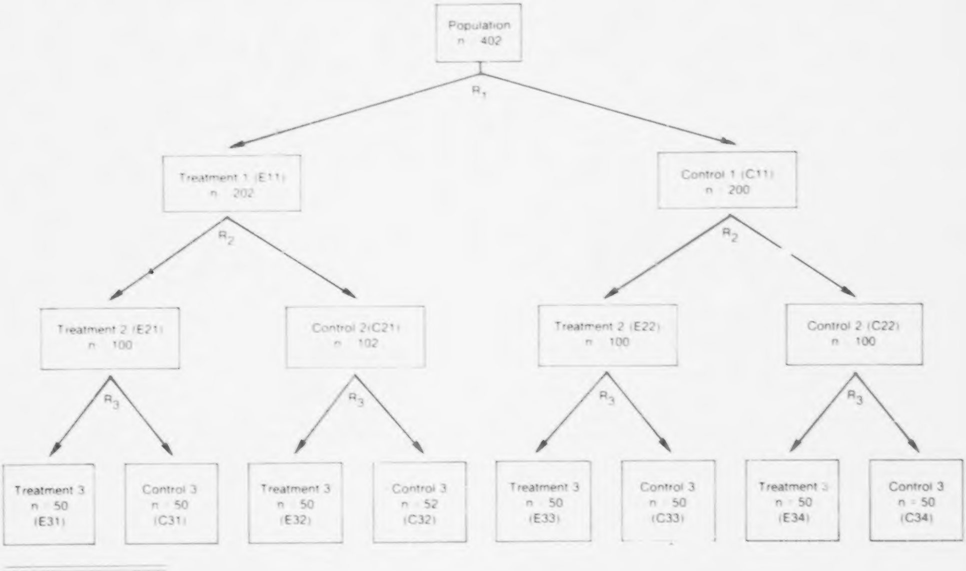
Reprint requests: Steven R. Eastaugh, Sc.D., Assistant Professor of Health Economics, Sloan Program of Hospital and Health Services Administration, Graduate School of Business and Public Administration, Malott Hall, Cornell University, Ithaca, New York 14853.

0025-7079/82/1000/1001/$01.65 © J. B. Lippincott Co.

large body of literature that cites many health education approaches shown to be effective in goal attainment,[1] but unproven from a cost-benefit standpoint. Second, the public cry for efficiency and the demand to conserve resources will necessitate methods of assigning patients to treatments most appropriate to their needs and to their capacities to benefit from them.[2] Third, as is the case with the control of all chronic diseases, adherence to medical regimen through voluntary education is considered a major problem.

The data base consisted of 402 hypertensive outpatients from Johns Hopkins Hospital, who were assigned randomly into three health education interventions: 1) a clarification of the doctor's instructions; 2) family member reinforcement; and 3) a series of three group meetings with other patients. The interventions were applied individually and in sequential combination (Table 1, Fig. 1). Data for application of the methods was derived from a postintervention field questionnaire.

The rationale for all three social support interventions that improve compliance has been reviewed elsewhere.[3,4] Some supporting evidence for the third intervention, group meetings designed to shift the patient's orientation, has been provided by Conte et al.[5] The rationale for the second intervention, family member reinforcement, has been provided by Fass et al.[6] and Levine et al.[7] Less costly forms of intervention, such as the provision of written information to patients, have proven to be ineffective in improving compliance.[8] Physician-directed programs to indirectly improve patient compliance by means of clinician education, even when linked with computer feedback, have proven to be equally disappointing.[9] Although there is an imperfect link between changes in physician attitude and changes in behavior, there is some evidence to suggest


Population
n = 402
R1
Treatment 1 (E11)
n = 202
Control 1 (C11)
n = 200
R2
R2
Treatment 2 (E21)
n = 100
Control 2 (C21)
n = 102
Treatment 2 (E22)
n = 100
Control 2 (C22)
n = 100
R3
R3
R3
R3
Treatment 3
n = 50
(E31)
Control 3
n = 50
(C31)
Treatment 3
n = 50
(E32)
Control 3
n = 52
(C32)
Treatment 3
n = 50
(E33)
Control 3
n = 50
(C33)
Treatment 3
n = 50
(E34)
Control 3
n = 50
(C34)

R_1, R_2 = Random assignment

R_3 = Stratified random assignment, i.e.—stratified on prior treatment

n = Sample size assigned

FIG. 1. Study Design: sequential factorial provision of three experimental interventions.

TABLE 1. Eight Study Groupings and Percentage of Patients with Blood Pressure Under Control

Study Status	Combinations Utilized*	Sample Size	Percentage of Sample With BP Under Control: Initial Baseline Measures	Percentage of Sample With BP Under Control: Posttest Following Interventions
E31	RD, FR, SG	50	38%	66%
C31	RD, FR	50	45%	55%
E32	RD, SG	50	40%	53%
C32	RD	52	40%	41%
E33	FR, SG	50	41%	54%
C33	FR	50	37%	48%
E34	SG	50	35%	52%
C34	None	50	41%	42%

* RD is intervention 1 (reinforcement of doctor's message at the exit interview), FR is intervention 2 (family, reinforcement therapy), and SG is intervention 3 (small group therapy).

that attitudes might be instilled more effectively in the last 2 years of medical school.[10]

Health education can be described as a voluntary method for producing change in individuals so that desired behavioral and, presumably, physiologic changes are accomplished. One important principle in health education is the need for continuous application of treatments for any long-term behavioral change.[11] The primary difficulty is the uncertainty of effectiveness and the decay of effect over time. Because of the tendency for a single intervention to decay, health education needs to be a continuous process. Green[12] and Kasl[13] have outlined a number of health education models and treatment interventions designed to affect predisposing, enabling, and reinforcing factors. Unfortunately, many health education efforts have proven to be ineffective.[14]

Background

Hypertension is a largely asymptomatic disorder that individuals tend to disregard and thus not treat. Its estimated prevalence for adults aged 18–79 ranges from 15 to 18 per cent.[15] The prevalence rates are twice as high for blacks as whites and the disease tends to be more severe for blacks. The probability of having hypertension also increases with age. For example about 4 per cent of all individuals in the age groups 17–24 years have hypertension and 35 per cent of all individuals in the age group 65 years of age and over have hypertension. An estimated 9 million Americans have used antihypertensive medication for the condition in the previous 6 months, but 60 per cent of these patients still have high blood pressure and definite hypertension. In summary, about 5 million Americans potentially could be aided by measures that improve the rate of adherence to medication.[16]

For a long time hypertensive disease has been recognized as a serious public health problem that causes morbidity, disability, and mortality. A dilemma of treatment is that a majority of people who have hypertensive disease are symptom free. Even a slight elevation in blood pressure levels causes a marked increase in the risk of major cardiovascular complications. For example, the Society of Actuaries 1959 study showed that 35-year-old men with blood pressure readings higher than 140/90 mm Hg, without concomitant impairments had a 19.4 per cent mortality rate over the next 20 years compared with a 11.0 per cent rate for nonhypertensives.[17] For all causes of deaths, the ten-year mortality rate for men with diastolic pressure 95 to 105 mm Hg was 60 per cent higher than the rate for normotensives. The mortality rate was 200 per cent higher for those with diastolic blood pressure of 105 mm Hg and greater.

Many studies have demonstrated the effectiveness of antihypertension medication.[17, 18] The most publicized studies were based on data from the Veterans Administration Cooperative Study Group on Antihypertensive Agents.[20] For a carefully selected population, this prospective study of drug-treated and control-group hypertensives utilizing a double-blind randomized protocol showed that medications will reduce the probability of morbidity and mortality. Increased benefits were a function of the following risk factors: age, race, sex, target organ damage, family history, hyperglycemia, and hypercholesteremia. The 210 VA study patients with moderate hypertension (105–114 mm Hg) had a 3:1 difference in major complications between the treatment and control groups during the 3-year follow-up period. The treatment was efficacious against stroke, congestive heart failure, accelerated hypertension, and progressive elevation of blood pressure, but not against myocardial infarction.[21]

For the VA patients with moderately severe and severe hypertension, 21 out of 70 control patients developed major complications during 18 months of follow-up, but only 1 out of 73 treated patients developed a complication. The treatment given consisted of hydrochlorothiazide, reserpine, and hydralazine hydrochloride. After 2.4 months of active treatment, an average reduction of 42 mm Hg systolic and 29.7 mm Hg diastolic occurred for moderately severe and severe cases. The placebo treatment group showed no significant changes in average blood pressure levels.

At the present time improved adherence to treatment is a greater policy problem than the need to screen populations.[18, 22] Many individuals are aware that they have hypertensive disease and have started treatment, but either do not comply to medical instructions or drop out of treatment once they have started. It would not be benefical to screen these patients again.[22] The cost-effectiveness of a nationwide screening program needs careful scrutiny. In this context, Junod and Gutzwiller[23] have designed a medical and public health program evaluation project to pinpoint the most cost-effective modes of reducing cardiovascular disease risk factors. A major component of their cost-effectiveness model is based upon the adherence to medication rate for detected patients.

According to the literature, education, income, occupation, and race are significant socioeconomic factors in determining a person's ability to follow an antihypertensive program.[24] Other characteristics, such as Wallston's scale for locus of control, also have been identified as significant factors that have an impact on the treatment of hypertensives. Naditch found that a perceived external locus of control and a high level of discontent is directly related to having hypertension.[25, 25a]

Mental health variables relating compliance behavior or receptivity to health education interventions have not been studied extensively. Bloom and Monterossa,[26] Simpson,[27] and Garcia-Webb[28] considered the relationship between depression and hypertension and found that the treatment of one may induce the occurrence of the other. In several prospective studies depression has been found useful for prediction. Nadelson et al. found that three different measures of depression could differentially predict functional as opposed to organic disorders with a 90 to 95 per cent correct classification rate.[29]

Design

The three interventions were given in a sequential factorial experimental design.

This allows each intervention or any combination of the three interventions to be evaluated. The first intervention (treatment 1, Fig. 1) concerned clarification of the prescribed regimen, the second (treatment 2) attached social support by intervening with a family member, and the third (treatment 3) attempted to increase self confidence in compliance by changing the locus of control from an external orientation to an internal orientation. The sequential combination of interventions allowed each treatment or combination of treatments to be evaluated against a control group.

In a 2-month period, 402 patients that were identified as hypertensive from the General Medical and Hypertension Clinics of the Johns Hopkins Hospital were selected for the study. All the patients were outpatients that had been in treatment for at least 6 months. During the randomization process the experimental group was given the first intervention. The treatment 1 session was a reteaching and reinforcement of the medical practitioner encounter that immediately followed the medical intervention. All patients were identified in the waiting room and asked if they could stay 15 to 20 minutes after seeing the doctor and/or nurse to discuss their treatment for hypertension. If the patient was assigned to the control group, he or she was informed that no one was available at this time, but that they may be contacted in the future. Of the 407 patients eligible for the study, only five refused to participate, 202 received treatment 1.

Two months after the first intervention was initiated half of the control and experimental patients (200) were selected for interviewing in their homes. The average interview lasted 53 minutes. At that time, the name of a family member and permission to contact that person was obtained. The family member was selected by the patient based upon the question, "Who is the adult person in your household whom you talk with the most?" If the patient identified more than one person then he or she was asked with whom they spend the most time.

The second intervention (treatment 2) was an outreach effort in the patient's home. The family members were visited 3 months after the initial interview. They were given specific information on hypertension, what the patient should do, and special advice on what they could do to help. The sessions lasted an average of 30 minutes. Fass et al.[6] has described the intervention used in this study.

The third intervention (treatment 3) consisted of three 1.5 hour group awareness sessions. The purpose of this intervention was to increase the patients' self-confidence about controlling their own health and to change an individual from an external locus of control to an internal locus of control. Individual support was provided through the group discussion. Random sets of 16 patients were recruited for each set of group sessions. The sessions contained role playing, behavioral rehearsal, problem clarification, and general discussion of management and compliance in areas where they lacked self-confidence about their problems with hypertension. Bowler has discussed the rationale for the intervention and content of each session in detail.[30] Of the 200 patients selected, only 24 refused to participate. The sessions at Johns Hopkins Hospital were arranged to meet the patient's convenience and free transportation was provided.

The average age of the sample population was 53.7 years, 75 per cent female, and 93 per cent black. The average monthly income was about $330.00. The average education was between eighth and ninth grade. Twenty-three per cent of the population lived alone and fifty-five per cent had the cost of their medication covered by medical assistance. The effectiveness of the randomization of patients into unbiased groups was tested according to differences in sex, race, age, educational level, occupation, income, medication compliance (Table 1), and medication being paid for by medical assistance. No significant differences were found between the groups for selected baseline socioeconomic and medical variables after performing a chi-square test.

Methods

The decision process consisted of two steps. First, the variables exhibiting strong discriminating properties associated with participation in a health education intervention were selected mathematically. The results are presented in Tables 2–4. Discriminant analysis was performed to separate patients into high and low compliers within treatment groups. These results are available from the authors, but are not presented here because they essentially are in agreement with the compliance likelihood ratio tests presented in step 2 of the model. Depression, as measured by a self-report scale, was the first variable to enter the discriminant analysis, indicating that it contributed the most to separation of the groups.

In the second step, the likelihood of success (medication adherence) for experimental versus control groups (no health education effort) was developed for each of these selected discriminator variables and tested for their statistical significance. The 29 variables under consideration in step 2 are listed in Table 5. The proportion of compliers with a certain attribute was compared for the experimental versus control group. This ratio of proportions was formed and tested statistically. This ratio gives a direct estimate of the likelihood of success for one treatment versus another treatment. A likelihood ratio greater than 1 indicates that the health education intervention was effective in stimulating increased adherence to treatment. This ap-

TABLE 2. Independent Variables that Delineate Low and High Compliance in the Discriminant Function Analysis

Variables	Function	Related to
Age	−0.22	High compliance
Live with someone	0.25	Low compliance
Speak to someone for advice before the doctor	0.24	Low compliance
External locus of control	−0.43	High compliance
Luck is unimportant	0.21	Low compliance
Stay in hospital for HBP	0.14	Low compliance
Health is poor (perception)	0.17	Low compliance
Not worried about health	0.25	Low compliance
Control over life	0.17	Low compliance
High patient perception of staff's expectations for appointment keeping	−0.23	High compliance
Long clinic waiting time	0.40	Low compliance
Understood doctor/nurse	0.28	Low compliance
Believe that medicine works for HBP	−0.24	High compliance
Belief that HBP is serious	0.25	Low compliance
Belief in seriousness of your condition	−0.36	High compliance
Phase II—control C2, Figure 1)	0.22	Low compliance

Low compliance: Centroids, 0.43; N = 106. High Compliance: Centroids, −0.59; N = 77.
Recognition Classification Rate: 75% (for all these stable variables).
Canonical Correlation: 0.51.

TABLE 3. Variables That Indicate Participation in Family Reinforcement (Treatment 2)

Variables	Function C	Related to
Phase I—Control	−0.69	Nonparticipation
Dissatisfaction	+0.58	Participation
*Someone for advice before doctor	+0.27	Participation
Dreams about HBP complications	−0.51	Nonparticipation
Understood doctor/nurse	+0.24	Participation
Depression	+0.45	Participation
Influence over things	+0.22	Participation
*Control over life	−0.18	Nonparticipation
Family does not help follow doctor's advice	−0.28	Nonparticipation
*Health is poor (perception)	−0.26	Nonparticipation
Sex—female	+0.22	Participation
*Race—white	Positive	Participation
*Medical assistance does not cover medication cost	Positive	Participation
*Education	Negative	Nonparticipation

Participation: Centroids, +0.687; N = 35. Nonparticipation: Centroids, −0.617, N = 39.
Recognition Classification Rate: 81% (for all variables).
Recognition Classification Rate: 68% (for stable variables only).
Canonical Correlation: 0.66.
* Discriminating variables when only stable variables are considered.

proach is analogous to the criteria used in sequential decision making (Flagle[31]).

Linear discriminant functions are effective for both prediction and description of decision processes. Discriminant functions recently have been applied to predict medical problems. Thomas and Greenstreet used discriminant analysis in studying psychobiological characteristics in youth as predictors of five disease states: suicide, mental illness, hypertension, heart disease, and cancer.[32] After the success of decision modeling in the area of diagnosis classification,[33] it would seem logical to extend decision modeling into treatment assignment.

The discriminant analysis has several steps. First, variables are selected for inclusion in the analysis based upon a criterion that deals with group separation, in our study, the criterion maximizing the square root of the distance between the two closest groups.[34] Before a variable is selected it has to meet another criterion, that of increasing total group separation. The first variable selected has the highest value on the criterion. Then the single variable is paired with each of the other available variables, one at a time, and the selected criterion computed. The new variable, in conjunction with the old variable, which has the highest value on the criterion, is selected. The process is continued until no improvement is made or a minimal level is reached, e.g., the variable is considered for selection if its partial multivariate F ratio is larger than a specific value. At the beginning of each variable selection step, each previously selected variable is checked to see if it still contributes to the discrimination. If a variable loses its discriminating power, it is removed, e.g., the multivariate F ratio is less than a certain value.

TABLE 4. Variables That Indicate Participation in Small Group Meeting (Treatment 3)

Variables	Function C	Related to
*Age	+0.56	Participation
Family does not help follow doctor's advice	+0.43	Participation
*Dreams about HBP complications	+0.41	Participation
Belief in benefit	+0.41	Participation
Phase I—Control (C1, Figure 1)	−0.32	Nonparticipation
Belief in seriousness	+0.52	Nonparticipation
Understood doctor/nurse	+0.35	Participation
*Patient worried about HBP	−0.24	Nonparticipation
Race—white	−0.35	Nonparticipation
Medicines work for HBP	+0.39	Participation
Dissatisfaction	+0.34	Participation
High blood pressure is serious	+0.37	Participation
Influence over things	+0.20	Participation
*Education	+0.26	Participation
Live with someone	+0.17	Participation
*Patient's perception of staff's expectations of patient's appointment keeping (high)	Negative	Nonparticipation

Participation: Centroids, +0.531; N = 58. Nonparticipation: Centroids, −0.670; N = 46.
Recognition Classification Rate: 75% (for all variables).
Recognition Classification Rate: 67% (for stable variables only).
Canonical Correlation: 0.60.
* Discriminating variables when only stable variables are considered.

After the important variables are selected, the corresponding matrix is solved to find a weighted function of the variables such that the F ratio for the weighted (w) mean scores (x) for the samples in a one-way analysis of variance is maximized. The problem can be formulated in the matrix algebra as the set of weights that maximizes the expression:

$$\lambda = (w'\ W\ s)/(w'\ B\ w),$$

where B = between groups' p × p matrix of sums of squares and cross-products of the x scores about their respective means,

W = the corresponding within group matrix,

and the p × 1 column of weights is represented by vector w (with transpose w'). Since the expression involves scalar quantities and, if the total size is N, the degrees of freedom for "between" and "within" groups are (K-l) and (N-K), respectively,

$$\lambda = \frac{K - 1}{N - K} F.$$

Therefore, maximizing λ is equivalent to maximizing the F-ratio between groups and thus maximizing the difference between group means.

The first function in the process will explain the largest share of explained variance, by definition. The relative contribution of each variable to the respective function can be obtained from the weights for the standardized discriminant function. In the standardized function, over all cases, the scores for one function will have a mean of 0 and a standard deviation of 1. If a weight has a value of 0.70 then the variable is twice as important as a variable with a 0.35 value. Nonstandardized coefficients do not represent relative importance because they have not been adjusted for the measurement scales and variability in the original variables.‡

Scales previously have been developed for measurement of the adherence to medication dependent variable and the six independent variables: depression, dissatisfaction, belief in seriousness, belief in benefits, health external locus of control, and patient's perception of staff's expectations for patient's appointment keeping. The reliability coefficient for measuring the internal consistency of the times used in the scale is Cronbach Alpha.[35] The previously published[7] adherence to medication scale had a Cronbach Alpha of 0.92 with four items, i.e., there is a high tendency toward consistency from one question to the next. The depression scale had a high reliability of 0.85 with seven items.

A complex construct, such as depression, reflects a broad range of signs and symptoms of the clinical depressive syndrome. Symptoms of dysphoric affect and mood, withdrawal of interest in life activities, lack of motivation, and loss of vital energy are all depressive symptoms. In addition, feelings of helplessness and futility also are key signs. The first six items in the depression scale come from the SCL-90 Hopkins Symptom Checklist scale reported by Derogatis.[36] The patient was asked to read a list of problems and complaints and "report how much discomfort that problem has caused during the past year" (feeling blue, lonely, hopeless, worthless, suicidal, or having no interest in things). A seventh item was included in the depression scale and targeted specifically to the hypertensive patients: "Do you have feelings of hopelessness about your high blood pressure?" §

Results

The compliance scale consists of four adherence to medication items: ever forget your medicine, careless at times about taking your medicine, sometimes stop taking your medicine if you feel better, and sometimes stop taking your medicine if you start to feel worse. Table 2 summarizes the results of the discriminant function analysis for high and low compliance. The results indicate that factors such as age, believing in the medication, staff expectation for keeping all appointments, and belief in the seriousness of hypertension all contributed to higher compliance behavior.

Tables 3 and 4 summarize the results of the discriminant function analysis of "participation in the family member reinforcement" and "participation in the group meeting," respectively. In Table 3, the function measures "dissatisfaction" and "depression" on the positive (higher) end, i.e., indicating higher participation in the family reinforcement therapy (intervention 2). The positive scores relate to participation in the family-member reinforcement, the negative scores to nonparticipation. The recognition classification rate is the percentage of the cases that the model correctly classified based upon preassigned classifications. The recognition classification rate of 81 per cent is considerably over the 50 per cent that would be expected by chance.

---

‡ Another measure for importance of a function is the canonic correlation, which is derived from the eigenvalues (latent roots). The canonic correlation represents the ability of each function to separate groups by the per cent of variance contained in the discriminating variables that is explained by the function. The canonic correlation squared is the proportion of variance in the discriminant function explained by the groups. The centroids for each function are the mean of each group for that function. Comparison between centroids represents the separation of the groups.

§ The questionnaires used in the development of the scales are available from the authors upon written request. The depression scale has been validated against other self-report instruments and scales, such as the Minnesota Multiphasic Personality Inventory and the Zung inventory for measuring clinical depression,[37] but has yet to be validated against any depression scale based on paraprofessional interviews (e.g., the Hamilton inventory).[38]

TABLE 5. Exogeneous and Endogeneous Independent Variables

Exogeneous Variables (Stable and not affected by the intervention)	Endogeneous Variables (Variables that are dependent to some unknown extent on the intervention)
Sex—female	Family does not help follow doctor's advice
Race—white	Dissatisfaction with the medical care system scale
Age	Belief in seriousness scale
Working status	Believe that high blood pressure is serious
Education	Believe that medicines work for high blood pressure
Medical assistance does not cover drug cost	Don't know enough to care for HBP
Live with someone	Understood doctor/nurse
See someone for advice before doctor	Treatment 1
Perception of long clinic waiting time	Treatment 2
Health external locus of control scale	Treatment 3
Control over life (ladder scale)	
Influence over things	
Luck is unimportant	
Depression scale	
Stay in hospital for high blood pressure	
Dreams about HBP complications	
Patient worried about high blood pressure	
Health is poor (perception)	
Not worried about health	
Perception of staff's expectations of patient's appointment keeping (scale)	

The exogenous and endogenous independent variables considered in the analysis of adherence to medication are given in Table 5. Tables 6 and 7 contain the decision rules of the likelihood ratio test. The results from the analyses designed to discriminate compliers from noncompliers are utilized for determining the significant variables. We then compute likelihood ratios,[‖] and consider the interrelationships of the health education interventions being evaluated. Statistically significant results were found with the likelihoods developed for the 18 selected independent variables (Table 7). The consistency between the results of the discriminant analyses and the likelihood calculations corroborate the results.

The first intervention produced an inverse effect between compliance and depression (Table 6). The first intervention may be effective for individual's with no depression. It may be ineffective for those individuals with high depression and those individuals who are not worried about their health. The first intervention is a reinforcement of the doctor's instruction, which allows patients to take better care of themselves; this is an active medical reg-

---

[‖] A likelihood can be constructed from any two probabilities or proportions. The likelihood ratio of high compliance versus low compliance for a treatment may be interesting but it cannot be statistically tested for difference in proportions. For two proportions to be tested statistically with a binomial "t" test, they must be independent. Thus, for the likelihood ratios to meet this criteria, they must compare proportions from groups that are independent. This was accomplished in our study by the use of the randomized experimental design approach.

TABLE 6. Adherence to Medication Likelihood as a Function of Depression

Treatment Phase(s)	Given: High Depression*			Given: Medium Depression			Given: Nondepression		
	Likelihood of High Compliance†	Significance		Likelihood of high Compliance	Significance		Likelihood of High Compliance	Significance	
		t	p		t	p		t	p
Phase I E1 vs. C1	0.66	−1.32	<0.10	0.91	−0.52	—	1.52	1.32	<0.10
Phase II E22 vs. C22	2.50‡	2.54	<0.01	1.13	0.52	—	0.42		
Phase III E34 vs. C34	0.36			1.52	1.09	<0.20	0.25		
Phase I & II E21 vs. C21	0.73	−0.59	—	1.76‡	2.46	<0.01	1.86‡	1.83	<0.05
Phase II & III E33 vs. C33	0.90			0.89	−0.35	—	§		
Phase I & III E32 vs. C32	1.00			1.74	1.34	<0.10	1.61		
Phase I, II & III E31 vs. C31	0.88			0.93		0.25		0.68	

* No depression score is 7, a medium depression is greater than 7 to and including 16, a high depression is greater than 16 to and including 35.

† A low compliance is 5 to and including 13 and a high compliance is greater than 13 to and including 15. There were natural break points in the response distribution.

‡ Significant at the 5 per cent level. Sample sizes of high compliers in these three cases, reading left to right, are 16, 40, and 17, respectively.

§ All cases were high compliers.

TABLE 7. Likelihood of Success in Adherence to Medications As a Function of Type of Treatment

Interventions*	Variables	Likelihood*	Percentage Increase in High Compliers	Significance†	Actual Sample Size
Treatment 2					
E22 vs. C22	Depression (high)	2.46	156%	0.01	(11/17)/(5/19)
	Health is fair/poor	1.58	58%	0.05	(8/23)/(11/25)
Treatment 3					
E34 vs. C34	Education (9th grade or more)	0.48	−52%	0.05	(6/20)/(8/13)
	Health is excellent/good	0.47	−53%	0.05	(6/18)/(5/17)
Treatment 1 & 2					
E21 vs. C22	Medical assistance pays for medicine	1.54	54%	0.09	(24/40)/(16/41)
	Clinic waiting time (all right)	1.54	54%	0.05	(24/32)/(17/35)
	Age (19 to 53)	2.0	100%	0.01	(20/39)/(10/39)
	Not working	1.55	55%	0.05	(30/47)/(21/51)
	Not worried about health (little or not at all)	2.04	104%	0.002	(27/53)/(15/60)
	Health external LOC	1.83	83%	0.05	(16/35)/(7/23)
Treatment 2 & 3					
E33 vs. C34	Education (less than 9th grade)	0.48	−52%	0.05	(5/17)/(8/13)
	Education (9th grade or more)	2.13	1.95	0.05	(10/17)/(5/18)
	Says health is good/excellent	0.43	−57%	0.05	(5/16)/(5/7)
	Says health is fair/poor	1.79	79%	0.05	(11/20)/(8/26)
Treatment 1 & 3					
E32 vs. C34	Education (less than 9th grade)	0.39	−71%	0.05	(5/21)/(8/13)
	Education (9th grade or more)	1.97	97%	0.05	(12/22)/(5/18)
	Says health is excellent/good	0.43	−58%	0.05	(5/18)/(9/14)
Treatment 1, 2 & 3					
E31 vs. C34	Clinic waiting time (all right)	1.70	70%	0.05	(13/17)/(8/15)

* Refer to Figure 1

† All but one item in this column are significant at the 5 per cent level or better.

imen. One possible explanation is the state of "no depression" is related to being receptive to active medical regimens and the reverse is true for high levels of depression.

The second intervention was effective for patients with high self-reported levels of depression and the feeling that one's health is fair or poor (Table 7). Depression was not important in the discriminant analysis without the inclusion of the other variables. One rationale for explaining why depression is an important characteristic for indicating responsiveness to the second intervention is that a depressed person generally tends to be passive and should be influenced by family members who have learned how to reinforce the patient in the intervention.

The decision rules listed in Table 7 indicate that a combination of the first and second interventions are best, in general. For the practitioner this would mean that everyone should receive clarification of the doctor's message and family member reinforcement that can be tailored to the patient and his family member (treatment phase 1 and 2). There are exceptions to the superiority of treatments 1 and 2. The highly depressed patient does better with only reinforcement of the family member (treatment 2), which is consistent with the condition of depression and related passivity. Patients with more than a 9th grade education do better with intervention 3 combined with one or two. Surprisingly, combinations of interventions had a detrimental (decreased adherence) effect on patients with less than a 9th grade education.

The most important limitation in generalizing these results is the sample population. As the results are extended beyond this temporal point and population, the associated limitations must be considered. In a more technical vein, multicollinearity between patient characteristics critical for triaging patients is a limited problem considering the low Pearson correlation coefficients. Depression was seldom significantly related to "health is poor (perception)," "not worried about health," "control over life," and "age" (0.27, −0.38, −0.18, and −0.18 correlation coefficients, respectively). "Age" was significantly related to "education" and "live with someone" (correlation coefficients −0.46 and −0.24, respectively). Older people tend to live alone and have less education. There were four variables that were considered important in the design phase but were not analyzed because the response rate by patients was too low. These were: "occupational level," "income," "medicine cost is OK," and "visit cost is OK."

One final caveat is in order. If the sample size had been larger, a holdout sample could have been used to test the predictive ability of the descriptive analyses presented. Such a technique should be employed with data from the sample of 3,400 Australian hypertensives that were randomized into either therapy group or placebo.[39]

Weighing Costs and Benefits

A limited benefit–cost comparison was developed to evaluate the results of the triage methods. Benefits had to be defined more broadly than simply the employer benefits of hypertension health education projects accrued with a 5-year time period.[40, 41] The model compared the costs of the intervention(s) versus the total savings from the intervention(s). Two types of benefit, prevented (foregone) indirect earnings loss and direct medical costs, were considered for our study. The first benefit measure was the estimated preserved earnings per person having a diastolic blood pressure less than 94 mm Hg versus 95 mm Hg or greater. The second benefit measure was the savings in expected annual cost per person for treatment of cardiovascular disease and nephritis (excluding cost of hypertension treatment per se) for a diastolic blood pressure below 94 mm Hg versus a level above 95 mm Hg. The benefit-adherence ratio coefficient came from a 1975 study by Stokes and Carmichael.[42] The data was inflated to 1979 dollars by the medical price index component of the consumer price index and earning trend index.[43]

The cost figures were adjusted for age and sex in the study population with the exception that the figures for the population under 35 years of age (9.5 per cent) were applied to the 35–44 age data and the figures for population over 64 years of age (25.6 per cent) were applied to the 55–64 age data because the 1975 raw data were not available. This approach produced a conservative estimate of savings since the population over 64 years of age would have a higher associated savings. Problems in underestimating the benefits and costs of personal medical care[44] and health programs[45] have been reviewed in detail elsewhere. The other assumption was that being defined as a high complier (42.4 per cent of the study population) or a low complier (57.8 per cent of the study population) is perfectly correlated with blood pressure control. In analyzing the cost of sequelae, blood pressure control was defined as a diastolic between 85 mm Hg and 94 mm Hg, and not being in control was defined as a diastolic blood pressure between 95 mm Hg and 104 mm Hg.

The costs of the treatments were estimated by adding the marginal personnel cost to provide each intervention. The cost of each intervention annually was estimated to be $5.35 for the clarification of the doctor's message; $19.14 for the family member reinforcement; and $17.00 for the series of three group meetings. The annual savings in earnings loss per person for control versus noncontrol was $53.00. The savings in expected annual cost per person for treatment of cardiovascular disease and nephritis for control versus noncontrol was $56.00. The limited benefit–cost model for the health education triage methods including the cost of the health education interventions is:

$$R = \frac{IHC \times (S_1 + S_2)}{PC} =$$

Benefit to cost ratio for intervention programs.

IHC = increase in number of additional high compliers with treatment relative to the control group (per 100 patients).

S_1 = savings per high complier in earnings loss per person for coantrol group versus noncontrol group.

S_2 = savings per high complier in expected annual cost for treatment of cardiovascular disease and nephritis for control group versus noncontrol group.

PC = additional treatment cost for intervention program versus control group.

Given the assumptions, the cost and savings information, and the effectiveness of the intervention programs, the decision to triage is most cost-beneficial for the 24 per cent highly depressed population (defined in Table 6). The benefit–cost ratio of 1.24 for hypertension control for persons in the age range of 35–65 (Option 4, Table 8). In other words, selecting the 24 per cent highly depressed subpopulation and providing family member reinforcement to this group is more cost beneficial than giving everyone the special health education intervention. Previous studies demonstrating a cost–benefit ratio in the 1.1–1.3 range may not stand the test of time in claiming statistic significance above the 1.0 level when applied to a larger population, i.e., less susceptible to the Hawthorne effect.[46] In other words, the newness and small scale of the program and the extent to which the patients are aware that they are part of a research project, may make the degree of program effectiveness less

TABLE 8. Simple Benefit-To-Cost Comparisons of Triage Options Versus the Option Not to Triage

Option	Triage	Type of Patients Triaged	IHC*	Health Education Intervention(s)	Benefit Cost Ratio
1	Yes	High level of depression	65–26	Family reinforcement (FR)	2.20†
2	Yes	Medium level of depression	58–33	FR + message clarification	1.15
3	Yes	No depression	65–35	FR + message clarification	1.33
4	No	All patients	60–32	FR + message clarification	1.24

* IHC = increase in number of high compliers with treatment versus control per hundred patients.
† Only option 1 has a significantly better ratio for triaging in comparison to not triaging.

generalizable to future situations. These results must be considered preliminary because of the lack of long-term follow-up. A 5-year following and larger sample size also would strengthen confidence in predictive applications of the discriminate analysis.

Conclusions

The premise of this study is that patients will derive a greater benefit from a health care system that assigns them to a health education treatment that is more effective for them. This assumes that there are certain characteristics of the patient that predispose him to respond better to one treatment than another and that these characteristics are measureable.[47] The decision rule in our model (Table 7) is to assign patients into health education treatments based on likelihoods of success statistically significant at the 5 per cent level. For example, the characteristic depression, as measured by the response to seven questions, is found to be significant in explaining treatment compliance. The effectiveness of the family reinforcement intervention for increasing the adherence to medication is statistically effective for highly depressed patients. Considering this type of differential effectiveness for patients with various characteristics, the ability to assign patients into health education intervention has the potential to maximize the achievable benefits of hypertension detection efforts. Future research should consider whether the annual cost of these special interventions continues to yield the same annual benefits, diminishing benefits, or increased benefits.

More precise data linking adherence to symptom decline and to improved health would provide greater accuracy for policy makers. The actual relationship between our simple high and low dichotomy for compliance and diastolic blood pressure levels also should be established definitively. However, no matter how refined the model, many dynamics of the situation could not be considered given our present poor state of information concerning outcomes of hypertension treatment. Better epidemiologic studies linking symptom reduction (blood pressure decline) to better health are necessary.[48] The major finding of our analysis is the potential importance of one mental health variable in explaining the effectiveness of health education efforts. The results in the depression scale would encourage the inclusion of other standardized measures of mental health symptoms in future studies.

References

1. Green L. Determining the impact and effectiveness of health education as it relates to federal policy. Health Educ Monogr 1978;6:28.

2. Eastaugh S. Cost of elective surgery and utilization of ancillary services in teaching hospital. Health Serv Res 1979;14:290.

3. Green LW, Levine DM, Deeds S. Clinical trials of health education for hypertensive outpatients: design and baseline data. Prev Med 1975;4:417.

4. Glanz K, Kirscht J, Rosenstock I. Linking research and practive in patient education for hypertension. Med Care 1981;19:141.

5. Conte A, Brandzel M, Whitehead S. Group work with hypertensive patients. Am J Nurs 1974;5:910.

6. Fass, MF, Green LW, Levine DM. Design of an experimental intervention to increase family reinforcement in the home for patient compliance. Presented at the American Public Health Association, Chicago, November 1975.

7. Levine D, Green L, Deeds S, Chawlalo J, Russell R, Finlay J. Health education for hypertensive patients. JAMA 1979;241:1700.

8. Morris L, Halpern J. Effects of written drug information on patient knowledge and compliance: a literature review. Am J Public Health 1979;69:47.

9. Dickinson J, Warshaw G, Gehlbach S, Bobula J, Muhlbaier L, Parkerson G. Improving hypertension control: impact of computer feedback and physician education. Med Care 1981;19:843.

10. Eastaugh S. Teaching the principles of cost-effective clinical decision-making to medical students. Inquiry 1981;18:28.

11. Mitchell JH. Compliance with medical regimens: an annotated bibliography. Health Educ Monogr 1974;2:75.

12. Green L. Toward cost-benefit evaluations of health education: some concepts, methods, and examples. Health Educ Monogr 1974;2:34.

13. Kasl SV. The health education belief model and behavior related to chronic illness. Health Educ Monogr 1974;2:433.

14. Battistella R. Disease prevention and health education. In: Battistella R, Rundall T, eds. Health care policy in a changing environment. Berkeley: McCutchan, 1978:274.

15. U.S. Department of Health, Education and Welfare. Health—United States 1976–1977. HEW Publication (HRA) 77-1232, Government Printing Office, Washington, D.C., 1977.

16. Breslow L. Risk factor intervention for health maintenance. Science 1978;200:908.

17. Lew EE. High blood pressure, other risk factors and longevity: the insurance viewpoint. Am J Med 1973;55:281.

18. Weinstein M, Stason W. Hypertension: a policy perspective. Cambridge: Harvard University Press, 1976.

19. Wagner EH, Warner JT, Slome C. Medical care use and hypertension." Med Care 1980;18:1241.

20. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertensive II: results in patients with diastolic blood pressure averaging 90 through 114 mm Hg. JAMA 1970;213:1143.

21. Freis E. The clinical spectrum of essential hypertension. Arch Intern Med 1974;133:982.

22. Langfeld SBI. Hypertension: deficient care of the medically served. Ann Intern Med 1973;78:19.

23. Junod B, Gutzwiller F. Community based cardiovascular risk factor experiment. Medecine Sociale et Preventitive 1977;22:157.

24. Nelson EC, Stason WB, Neutra RR, Soloman HS, McArdle PJ. Impact of patient perceptions on compliance with treatment for hypertension. Med Care 1978;16:893.

25. Naditch MP. Locus of control, relative discontent and hypertension. Social Psychiatry 1974;9:111.

25a. Lau R, Ware J. Refinements in the measurement of health-specific locus of control beliefs. Med Care 1981;19:1147.

26. Bloom J, Monterossa S. Hypertension labeling and sense of well-being. Am J Public Health 1981;71:1228.

27. Simpson FO. Hypertension and depression and their treatment. Aust NZ J Psychiatry 1973;7:133.

28. Garcia-Webb P. Hypertension and depression. Aust NZJ Psychiatry 1973;7:298.

29. Nadelson T, Flatley M, Lipsitt D. Quantitative indicators of functional disease. Massachusetts Institute of Technology Technical Report ACP-19, Lincoln Laboratory Project Report Series, Lexington, Massachusetts, 1971.

30. Bowler M. Internality training sessions with hypertensive patients. Working paper, Baltimore, Md., Johns Hopkins Hospital, 1977.

31. Flagle CD. A decision theoretical comparison of three procedures of screening for a single disease. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol IV, Biology and Problems of Health, 1966. Lecam L, Neyman J, eds. Berkeley: University of California, 1966.

32. Thomas CB, Greenstreet RL. Psychobiological characteristics in youth as predictors of five disease states. Johns Hopkins Med J 1973;132:16.

33. Taylor TR, Shields S, Black R. Study of cost-conscious computer-assisted diagnosis in thyroid disease. Lancet 1972,2:79.

34. Maxwell AE. Multivariate analysis. In: Holland W. ed. Data handling in epidemiology. London: Oxford, 1970:149.

35. Stanley JC. Reliability. In: Thorndike RL, ed. Education measurement. Washington, DC: American Council on Education, 1971:356.

36. Derogatis L, Lipman R, Rickles K, Uhlenhuth E, Cori L. The Hopkins symptom checklist: a self-report symptom inventory. Behav Sci 1974;19:1.

37. Derogatis L, Rickels K, Rock A. The SCL-90 and the MMPI: validation of a new self-report scale. Br J Psychiatry 1979;128:280.

38. Freedman A, Kaplan H, Sadock B, ed. Modern synopsis of the comprehensive textbook of psychiatry—II. Baltimore: Williams and Wilkens, 1976:1012.

39. Australian Therapeutic Trial in Mild Hypertension. Report by the ATTMH management committee. Lancet 1980;1:1261.

40. Hannan E, Graham J. A cost-benefit study of a hypertension screening and treatment program at the work setting. Inquiry 1978;15:345.

41. Haurer J, Goodman J. Evaluation of performance and cost in a hypertension control program. Med Care 1980;18:485.

42. Stokes J, Carmichael DC. Cost benefit analysis of model hypertension control. National High Blood Pressure Education Program, National Heart and Lung Institute, National Institute of Health, Bethesda, Md., 1975:32.

43. United States Bureau of Labor Statistics. Monthly Labor Review 1979;102(12):71.

44. Eastaugh S. Medical economics and health finance 1981. Boston: Auburn, pp. 55.

45. Eastaugh S. Effectiveness of community-based hospital planning: some recent evidence. Applied Economics 1982;14:(in press).

46. Shortell S, Richardson W. Health program evaluation. St. Louis: CV Mosby, 1978:47.

47. Williamson JW. Assessing and improving health care outcomes: the health accounting approach to quality assurance. Cambridge, Mass.: Ballinger, 1978.

48. Szklo M. The epidemiologic basis for hypertension control. Journal of Public Health Policy 1980;1:312.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Social Networks, Beliefs, and the Decision to Seek Medical Care: An Analysis of Congruent and Incongruent Patterns

EMIL BERKANOVIC, PH.D., AND CAROL TELESKY, M.A.

This article focuses on the impact of congruent and incongruent patterns of symptom specific beliefs and social network advice on the decision to seek medical care for symptoms. Data from 769 individuals who reported symptoms in the Los Angeles Health Survey are used to examine this issue. Further, the role of more general health orientations and social network influences in the decision to seek care is examined under these congruent and incongruent conditions. It is concluded that the more general factors, which have little effect on the aggregate, may play a role in decision making when beliefs and advice are incongruent. Further, it is concluded that reports of beliefs and advice about specific symptoms that have been given after the decision has been made to seek care may be valid and not the result of retrospective reconstruction.

IN A RECENT ARTICLE in Medical Care,[1] we examined the degree to which need, social structural, organizational, social network, and social psychological variables could predict the decision to seek medical care for symptoms. In that analysis, the independent variables were categorized into five main groups:

1. Need factors, including the characteristics of the symptom under study and the individual's overall health status.
2. Social structural factors that might predispose or enable individuals to seek medical care when confronted by symptoms.
3. Organizational factors that might enable individuals to seek medical care.
4. Social network patterns and general health orientations that might predispose individuals to seek care.
5. Specific social network influences and personal beliefs about the symptoms under study.

In a hierarchic multiple regression analysis, this set of independent variables explained 57 per cent of the variance in the decision to seek medical care for the symptoms reported over a study year. Of the 57 per cent of the variance explained, however, 41 per cent was explained by symptom-specific network advice and beliefs. Further, of the remaining 16 per cent, 12 per cent was explained by the need variables. These data are summarized in Table 1.

Because of the distribution of variance we concluded that the analysis reported in Table 1 was difficult to interpret. Specifically, the failure of the structural variables, the social network patterns, and the

---

From the School of Public Health, University of California, Los Angeles, California.

Supported by Grant 5-R18-CA18451, awarded by the National Cancer Institute, Department of Health and Human Services. A preliminary version of this article was presented to the International Society of Applied Anthropology, April 15, 1981, Edinburgh.

0025-7079/82/1000/1018/$01.25 © J. B. Lippincott Co.

TABLE 1. Hierarchial Multiple Regression Analysis of the Decision to Seek Medical Care for Symptoms

Step	Variables	Standardized Regression Coefficients*
I. Indicators of need	Subjective health status	0.05
	Number of chronic health problems	0.12†
	Reported disability from symptom	0.28‡
	$r^2 = 0.116$	
II. Sociodemographics	Total family income	0.05
	Socioeconomic index	−0.03
	Race/ethnicity (white vs. other)	−0.01
	Marital status (now married vs. other)	−0.02
	Sex (female = 1, male = 0)	0.04
	Age	0.09§
	$r^2 = 0.125$	
III. Organization of care	Continuity of care (index)	0.08§
	$r^2 = 0.130$	
IV. Social network characteristics and health orientations	Amount of network contact	−0.10§
	Amount of network consulting	−0.08
	Network support	0.11§
	Network size—total number of members	0.00
	Network advice to see doctor	−0.01
	Acceptability	0.02
	Accessibility	0.02
	Seriousness of illness	−0.02
	Perceived efficacy of care	−0.01
	Susceptibility to illness	0.06
	Motivation—concern with health	0.05
	Concern with medical costs	−0.03
	Tendency to delay seeking care	−0.05
	$r^2 = 0.151$	
V. Symptom specific networks and beliefs	Amount of network contact (log)	0.02
	Distance of network members consulted	0.11†
	Perceived influence of network	0.03
	Size of network consulted	0.02
	Network advice to see doctor	0.15†
	Perceived likelihood of recurrence	−0.06§
	Perceived efficacy of care	0.51‡
	Ease of going to doctor	0.09‡
	Perceived seriousness of symptom (index)	0.17‡
	$r^2 = 0.568$	

* Regression coefficients shown in Table 5 are presented in hierarchic fashion. Each coefficient represents the amount of change in the dependent variable that is attributed to a 1 unit change in the standard score of the predictor variable, independently of all other variables examined in the same or prior steps.

† $p < 0.01$.

‡ $p < 0.001$.

§ $p < 0.05$.

general health orientations to account for much of the variance in the analysis raised two troublesome questions. First, these findings failed to support the notion that behavior is linked to both social structure and enduring cognitive orientations. Thus, although substantial variance is explained, these data provide little insight into how social structure and prior experience, as expressed through cognitive orientations, impact on responses to individual symptoms. Second, since the symptom-specific data were collected subsequent to the decision to seek medical care, the validity of these findings may be jeopardized because of retrospective reconstruction. The issue of retrospective reconstruction is especially important because it is unlikely that symptom responses can be studied prospectively in the general population. To do so would require collecting data from the moment the individual becomes aware of a symptom.

The purpose of this article is to explore these effects further. Specifically, this analysis focuses on the impact of congruent and incongruent patterns of symptom specific beliefs and social network advice in the decision to seek medical care for these symptoms. In addition to providing further insight into how these symptom-specific factors combine to affect the decision to seek care, this analysis provides an opportunity for examining how variables that appear to have little effect on the analysis reported in Table 1 impact this decision under one or another of the combination of beliefs and network advice. Further, by comparing the impact of congruent and incongruent patterns, it may be possible to make some inference regarding the extent to which retrospective reconstruction is a viable interpretation of the results in Table 1.

Sample

The data to be reported were collected during a 1-year panel study as part of the 1976–1977 Los Angeles Health Survey (LAHS). The sample design for the LAHS was a three-stage random probability sample of Los Angeles County that was developed by the Institute for Social Science Research (ISSR) at UCLA.[2] From October through December 1976, demographic and health data were obtained from 1,210 respondents during the initial face-to-face interviews. The initial interview then was followed by seven telephone reinterviews at approximately 6–8 week intervals and a final face-to-face interview at the end of the panel year. Thus, 1 year of symptom experience is being analyzed in relation to the background information collected during the first interview. Details of variable measurement and construction are reported in Berkanovic, Telesky, and Reeder[1] and do not need to be reviewed here.

For the 769 respondents who reported any symptoms during the 1-year study, a total 1,679 "most important" symptom episodes were reported, with individual respondents reporting from 1–17 symptoms. Multiple symptoms reported by the same person are not independent, which violates a key assumption of multiple regression analysis. Thus, while the symptom data base contains 1,679 observations, the sample size for significance testing should be 769. Accordingly, we use the symptoms as the unit of analysis, weighting the data inversely by the number of symptoms reported by the respondent. This results in conservative significance tests that compensate for the lack of independence of the symptoms. One of the unique features of this study is that the predictor variables can be linked with specific symptom responses, including use of health services. Therefore, when examining a particular symptom episode we know how the person perceived the symptom and which social network members were consulted regarding that symptom. This link between specific health beliefs, social network characteristics, and symptom responses is a central component of the analysis.

Two questions regarding what the person believed to be true about each symptom reported during the study year were asked: How serious did you think this condition was? When you realized you had this problem, how much good did you think a doctor could have done you? Individuals who said they had spoken to a friend or relative about a symptom were asked "What kind of advice did (. . .) give you about this health problem?" Answers ranged from "wait and see" to "get medical advice immediately." Not one respondent reported being advised against seeing a doctor. For purposes of the present analysis, the respondents reporting symptoms are divided into those who were advised by at least one person to see a physician and those who were not so advised. Thus, congruence occurs when the person's beliefs support seeking care and the person has been advised to do so, or when the person's beliefs do not support seeking medical care and the person has not been advised to do so. The other two possible combinations of these variables indicate incongruence between what the person believes and what he has been advised to do by his social network.

Results

Table 2 presents the percentage of individuals who saw a doctor for symptoms for each of the congruent and incongruent conditions: When the network advised seeing the doctor and the person believed that a doctor could help relieve the symptom, 90 per cent of those with symptoms actually went to see the physician. In contrast, when there was no network advice to see a doctor and this belief was negative only 8 per cent went to see the doctor. In the incongruent conditions, about a third of those who did not believe a doctor could do them some good but who were advised to see a doctor by their network went to the doctor. However, 69 per cent of those who believed the doctor could help relieve their symptoms but who were not advised to see a doctor went. As one would expect, it appears that behavior is largely consistent with congruent personal beliefs and network advice, that without the reinforcement of advice substantially fewer people act according to their beliefs and that network advice can lead to behavior believed to be ineffective.

TABLE 2. Per Cent of Respondents Initiating a First Doctor Visit for a Symptom by Perceived Efficacy of Care and Network Advice to Seek Care (n = number of respondents in each category)

Proportion of Network Members Advising Respondent to See Doctor for Symptom	Perceived Efficacy of Care for Symptom	
	High Perceived Efficacy	Low Perceived Efficacy
Some or all	90.5% (339)	33.5% (79)
None	68.6% (553)	8.5% (411)

Table 3 presents four, four-step hierarchic multiple regressions of the decision to seek medical care for symptoms classified according to each of the four congruent and incongruent conditions in Table 2. Looking first at the total amount of variance explained, it may be seen that while 13 per cent of the variance in the decision to seek medical care is accounted for in the two congruent conditions, and 15 per cent is accounted for when the belief exists that the doctor can help relieve the symptom but the network does not advise a visit, 39 per cent of the variance in the decision to seek medical care is accounted for when the network advises seeing a physician but the person does not believe a doctor can do much to help relieve the symptom. While roughly equivalent amounts of variance in each of these analyses are accounted for in the first three steps, the fourth step, which enters general network characteristics and general health orientations, contributes substantially more to the variance explained

TABLE 3. Standardized Regression Coefficients for the Decision to Seek Medical Care for Symptoms by Network Advice and Perceived Efficacy of Care

Step	Received Advice/ High Efficacy (n = 339)	No Advice/ Low Efficacy (n = 411)	No Advice/ High Efficacy (n = 553)	Received Advice/ Low Efficacy (n = 79)
I. Indicators of need				
Number of chronic conditions	0.00	0.20*	0.16*	0.05
Disability from symptom	0.14†	0.09‡	0.21*	0.25†
Subjective health status	−0.09	0.05	0.01	−0.06
r^2 for Step I	0.02	0.06	0.08	0.05
II. Sociodemographics				
Total family income	−0.12	0.02	0.04	0.26†
Socioeconomic index	0.06	−0.07	−0.02	−0.05
Race/ethnicity (white vs. other)	0.14†	0.04	0.05	0.05
Marital status (now married vs. other)	0.03	−0.04	−0.11†	−0.05
Sex (female = 1, male = 0)	−0.08	0.03	0.02	0.08
Age	0.06	0.15*	0.09	0.08
r^2 for Step II	0.06	0.09	0.10	0.12
III. Organization of care				
Continuity of care	0.10	0.02	−0.04	−0.01
r^2 for Step III	0.07	0.09	0.10	0.12
IV. Social Network Characteristics and Health Orientations				
Amount of network contact	−0.09	−0.12	−0.13†	0.03
Amount of network consulting	−0.06	−0.07	−0.10	−0.06
Network support	0.05	0.14†	0.13†	0.01
Network size—total number of members	0.02	−0.02	0.02	−0.31†
Network advice to see doctor	−0.02	−0.11†	0.02	−0.07
Acceptability	0.08	0.03	0.00	−0.08
Accessibility	0.14†	−0.01	0.02	0.01
Seriousness of illness	0.04	−0.05	0.12†	0.16
Perceived efficacy of care	−0.11	0.03	0.04	−0.12
Susceptibility to illness	0.01	0.07	0.00	0.24‡
Motivation—concern with health	0.10	0.02	0.05	−0.03
Concern with medical costs	−0.06	−0.07	−0.12†	0.25‡
Tendency to delay seeking care	−0.03	−0.02	−0.06	0.34†
Total r^2 =	0.13	0.13	0.15	0.39
n	(339)	(411)	(553)	(79)

* $p < 0.01$.
† $p < 0.05$.
‡ $p < 0.10$.

among those who do not think a doctor can do much to help relieve their symptoms, but who are advised to go by their network. Further, the factors that are statistically significant at this step are network size, in which smaller networks seem to have more influence, perceived susceptibility to illness in general, concern about the cost of care, and self-reported tendency to delay seeking treatment. The only other variables to achieve statistical significance in this analysis are disability due to the symptom and income. Thus, among those who do not believe a doctor can do much to relieve their symptoms, but who are advised by their social networks to see a physician, characteristics such as greater disability, higher income, greater perception of one's susceptibility, a smaller social network, greater concern about the costs of care, and greater recognition that one tends to delay seeking medical attention for all symptoms, all contribute to the greater likelihood that care will be sought for any particular symptom.

Although the Betas for several variables are statistically significant in the other three analyses, it should be observed that, with two exceptions, none of them amounts to 0.20. As a consequence, none of the variables that are statistically significant in these analyses could be described as contributing substantially to the prediction of the decision to seek medical care. The two exceptions are need variables.

These analyses were replicated using an index combining belief that the symptom was serious and belief that a physician would regard the symptom as serious, in place of the belief that a doctor could help relieve the symptom. Table 4 presents the percentage of individuals with symptoms who saw a doctor in each of the congruent and incongruent conditions. The overall pattern in Table 4 corresponds to that in Table 2. Aside from slight differences in percentages, the only real difference between Table 4 and Table 2 is that network advice seems to be somewhat more effective in precipitating a physician visit when the individual does not believe the condition is serious, than when the person does not believe a physician can help. On the other hand, about as high a percentage of people who feel the symptom is serious seek medical attention without network advice as do those who believe the doctor can do them some good.

TABLE 4. Per Cent of Respondents Initiating a First Doctor Visit for a Symptom by Network Advice and Perceived Seriousness of Symptom (n = number of respondents in each category)

Proportion of Network Members Advising Respondent to See Doctor for Symptom	Symptom Serious	Symptom Not Serious
Some or all	85.6% (259)	49.3% (74)
None	65.2% (383)	13.4% (446)

Table 5 presents hierarchic multiple regression analyses for each of the congruent and incongruent conditions similar to those presented in Table 3. Again, the pattern and magnitudes of the total amounts of variance explained in these four analyses are quite similar to that found in Table 3. Thus, while less than 15 per cent of the variance is explained in the congruent conditions and in the condition when the symptom is believed to be serious but there is no network advice to go, 43 per cent of the variance is explained in the condition when the symptom is not believed to be serious but the network advises seeking medical attention. In the present analysis, substantial amounts of the surplus variance explained in this condition are accounted for with the sociodemographics and network characteristics, and the general health orientation items. Thus, in addition to income, marital

TABLE 5. Standardized Regression Coefficients for the Decision to Seek Medical Care for Symptoms by Network Advice and Perceived Seriousness of the Symptom

Step	Received Advice/ Serious (n = 259)	No Advice/ Not Serious (n = 446)	No Advice/ Serious (n = 383)	Received Advice/ Not Serious (n = 74)
I. Indicators of need				
Number of chronic conditions	0.06	0.11*	0.06	−0.07
Disability from symptom	0.13*	0.12†	0.09	−0.30‡
Subjective health status	0.17*	0.03	0.03	0.06
r^2 for Step I	0.03	0.03	0.02	0.09
II. Sociodemographics				
Total family income	−0.13	0.06	0.11	0.26†
Socioeconomic index	0.07	−0.07	−0.07	0.12
Race/ethnicity (white vs. other)	0.08	0.04	−0.02	0.00
Marital status (now married vs. other)	0.15*	−0.06	−0.08	−0.23‡
Sex (female = 1, male = 0)	−0.02	−0.09	−0.04	0.18
Age	0.09	0.08	0.18	0.37†
r^2 for Step II	0.07	0.05	0.05	0.28
III. Organization of care				
Continuity of care	0.17	0.07	0.10	0.09
r^2 for Step III	0.09	0.06	0.06	0.29
IV. Social network characteristics and health orientations				
Amount of network contact	0.02	−0.09	−0.14	−0.22
Amount of network consulting	−0.03	−0.06	−0.15*	−0.20
Network support	0.03	0.09	0.18*	0.36*
Network size—total number of members	0.06	0.06	−0.06	−0.23
Network advice to see doctor	−0.11	−0.10*	0.03	0.05
Acceptability	0.01	−0.03	0.01	−0.03
Accessibility	0.07	0.01	−0.01	0.03
Seriousness of illness	0.07	−0.11	−0.07	0.15
Perceived efficacy of care	−0.04	−0.03	−0.09	−0.01
Susceptibility to illness	−0.01	−0.01	0.05	0.02
Motivation—concern with health	0.05	0.05	0.03	0.02
Concern with medical costs	0.01	−0.06	−0.10	0.28‡
Tendency to delay seeking care	0.08	−0.09	−0.03	0.25‡
Total r^2 =	0.13	0.10	0.11	0.43
n	(259)	(446)	(383)	(74)

* $p < 0.05$.
† $p < 0.01$.
‡ $p < 0.10$.

status and age were also factors, the unmarried and older people being more likely to seek medical care on the advice of their networks but in opposition to their beliefs. Further, those who described their network as supportive, who were concerned about the costs of care, and who described themselves as delaying as long as possible before seeing a physician about an illness were more likely to seek medical attention in this condition. It also should be noted that, while disability was positively related to seeking medical attention when the individual did not believe a doctor could help, it is negatively related to seeking care when the person believes the condition is not serious. This suggests that restriction of activity may take place instead of a doctor visit when it is believed that the condition is not serious, but that activity restriction may precipitate a doctor visit when one's social network contradicts one's belief that a doctor can be of no help. Finally, it can be seen that although there are a number of statistically significant Betas in the regression analyses of the other three conditions, none of them achieves 0.20. Thus, none of the variables in these analyses are strongly related to the decision to seek medical care for symptoms in either of the congruent conditions nor are they strongly related to the decision to seek medical care when the individual believes the condition is serious but has not been advised to seek medical attention by his social network.

Conclusions

The first conclusion we drew from Table 1 is that the general characteristics of social networks and general health orientations have little effect on this decision. The present analyses suggest that these variables may influence the decision to seek medical care when they occur in interaction with the incongruent pattern of network advice to seek medical care and beliefs that are contradictory of that advice. Thus, the contribution of these variables to the amount of variance explained is 0.27 when the individual does not believe a doctor can help relieve the symptom and 0.14 when he believes the symptom is not serious. Because these variables make no appreciable contribution to the other analyses, these findings indicate that more general considerations are most likely to affect behavior when the opinions of members of a person's social network contradict what the person believes. Under these circumstances, behavior appears to be more amenable to the influences of general cognitive orientations. Indeed, this finding indicates that more enduring general factors may come into play in shaping behavior only when there is a conflict in the more specific and immediate factors relevant to the behavior. That similar influences did not appear in the other incongruent condition is explained by the fact that many members of the network may simply not advise the person on whether or not to seek care. Thus, for many of these people there may have been no conflict.

The extent to which retrospective reconstruction might be a problem also could not be determined from the data presented in Table 1. The analyses of congruent and incongruent patterns of network advice and personal beliefs, however, are at a minimum, supportive of a straightforward interpretation of Table 1. Thus, the two incongruent conditions, those in which retrospective reconstruction, to the extent that it occurred, did not lead to a consistent recall of the past, displayed patterns similar to what one might have expected were the reports of beliefs and network advice taken at face value. In the first incongruent condition, about two-thirds of those whose beliefs supported seeking medical care did so without network advice to seek attention. Nevertheless, when both beliefs and network advice support seeking care, a substantially larger percentage of the symptomatic go to the doctor. Thus, social network advice reinforces the probability that a behavior will occur as one might predict from knowledge about the beliefs related to that behavior.

In the second incongruent condition, when the network advises seeking care but beliefs do not support this behavior, fewer individuals actually seek care than in the other incongruent condition. Further, those individuals who do not believe the doctor can do much to relieve the symptom are less likely to respond to network advice to go to see the doctor than those individuals who do not believe the symptom is serious. This is precisely what one would expect, based on the social psychological tenet that a specific behavior is more readily predicted by beliefs that are directly related to that behavior than by beliefs that are tangential.[3] In this instance, the effects of network advice are greater when the belief with which the advice is incongruent is related less directly to the behavior being influenced. That is, individuals are more readily persuaded by networks to see a doctor when they believe the symptom is not serious than when they believe the doctor cannot help.

These findings appear to support the argument that self-reports regarding social network advice and personal beliefs about symptoms are valid. Indeed, if the self-reports were the result of retrospective reconstruction one would expect respondents to recall only congruent patterns of advice and beliefs. However, a large proportion of the respondents reported incongruent patterns. The conclusion that these self-reports of beliefs and advice regarding the symptoms are not the result of retrospective reconstruction, tentative though it is, has implications for research into the role of personal beliefs and network influences in the precipitation of decisions in a free-living population. As previously noted, it is extremely unlikely that prospective studies of such decision making will be possible in most circumstances. A prospective study of the decision to seek medical care for symptoms would require data collection from the moment the symptom is first experienced until the time when the decision regarding a visit to the doctor has been taken. Obviously, it simply is not feasible to collect data in this manner. The present evidence suggests, however, that data collected retrospectively may validly reflect the impact of both beliefs and network influences on the decision-making process.

References

1. Berkanovic E, Telesky C, Reeder S. Structural and social psychological factors in the decision to seek medical care for symptoms. Med Care, 1981, 19:693.

2. Sumner J. The 1976 LAMAS frame and master sample: technical description. Los Angeles: Institute for Social Science Research, 1976.

3. Fishbein M, Ajzen I. Belief, attitude, intention and behavior. Reading, Mass.: Addison Wesley, 1975.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Assessing the Effectiveness of Neonatal Intensive Care

PETER P. BUDETTI, M.D., J.D.,* AND PEGGY MCMANUS, M.H.S.†

Although neonatal intensive care has been praised widely for individual successes, its effectiveness has not been established systematically or conclusively. The literature consists principally of reports from individual intensive care units with sample sizes too small for statistical validation or generalization. This study analyzes the results obtained by recalculating and pooling the isolated reports. In addition, the findings of the solitary clinical trial, some scattered epidemiologic data, and several analyses of regional birthweight-specific time series data are reviewed. Taken together, this constitutes a body of evidence that supports the conclusion that intensive neonatal medical care has played a significant role in bringing about the impressive reduction in infant mortality that has taken place in this country since 1965.

NEONATAL INTENSIVE CARE is one of the factors often credited with the recent decline in infant mortality in the United States. The widespread diffusion and growing costs of newborn intensive care, together with concerns about the health of survivors and the potentially dehumanizing way such care may be delivered, however, have resulted in increased critical scrutiny of the benefits of neonatal intensive care. Policymakers, health professionals, and the public are raising serious questions about the costs of and outcome produced by highly technologic care for newborns. The specific issue to be resolved is whether or not increasing expenditures for newborn intensive care have resulted in measurable improvements in the health status of newborns.

This paper addresses what may be considered the principal indicator of the effectiveness of neonatal intensive care: whether premature and ill newborn infants do in fact have a better chance of survival with intensive care than without it. Randomized clinical trials would provide the most direct evidence of effectiveness, but few have been conducted. Instead the literature principally consists of epidemiologic surveys, reports from individual intensive care services, and birthweight-adjusted mortality rate analyses. Nearly all such reports make strong statements on behalf of medical care based on the remarkable declines in mortality during the years when neonatal intensive care proliferated.

Given the type of data available, however, some caution is necessary in inter-

---

* Institute for Health Policy Studies and Department of Pediatrics, School of Medicine, University of California, San Francisco, California.

† Institute for Health Policy Studies, School of Medicine, University of California, San Francisco, California.

Supported by Grant HS02975 from the National Center for Health Services Research. Based upon work done for the Congressional Office of Technology Assessment under Contract 933-22-60.0.

Reprint requests: Peter P. Budetti, Institute for Health Policy Studies, 1326 Third Avenue, San Francisco, CA 94143.

0025-7079/82/1000/1027/$01.45 © J. B. Lippincott Co.

pretation. Improvements in survival concurrent with advancements in intensive newborn care do not provide the information needed to determine the degree to which the gains in survival actually are due to intensive medical care of the newborn. Other factors such as lower prematurity rates or improved maternal health, nutrition, and intensive intrapartum care all undoubtedly have played a role.

The purpose of this article is to analyze the diverse and unsystematic literature as a step toward partitioning out the effect of neonatal care from that of other factors in determining newborn survival. This analysis addresses only survival and not the many questions about the functional status of survivors. Even this limited question is not resolved easily. On balance the available evidence does support the conclusion that neonatal intensive care has played a major role in improving the chances of survival of many newborns, particularly those of very low birthweight, but it remains impossible to quantify precisely how much of the improvement is due to medical care.

History of Neonatal Intensive Care and Infant Mortality in the United States

A historic perspective is needed to understand the debate over the effectiveness of modern methods of caring for newborn infants. In the early part of this century, most sick newborns died within the first few hours of life. Premature newborns were not expected to live more than a few days. In 1878, Dr. E. S. Tarnier instituted the use of a "warming chamber"—the first incubator—and was able to increase the chance of survival for a large number of premature newborns.[1] Similar machinery was widely used in international expositions, and later became common at amusement park sideshows featuring "incubator babies." Eventually incubators gained acceptance in the medical community and became standard equipment in most hospitals that had maternity services. As the survival period for premature infants lengthened, problems of nutrition and disease, to which such infants are prone, became evident and research on these problems began.

Facilities and medical techniques for care of the newborn have progressed rapidly since 1965 with the evolution of perinatal medicine and the development of associated medical technology. Care provided the newborn has passed through many phases and now involves the use of highly technologic, diagnostic, and therapeutic techniques and life-support systems. Along with changes in neonatal intensive care came changes in training physicians and nurses who deliver such care. Before 1960 responsibility for the care of the newborn was shared among general pediatricians, obstetricians, and general practitioners. In the early to mid-1960s, pediatricians with special interest in caring for sick newborns began to specialize in early infant care, leading to the development of neonatology as a pediatric subspecialty.[2] Now nurse clinicians and nurse specialists with advanced training and skills are responsible for delivering much of the intensive newborn care provided.

Neonatal intensive care typically is provided in organized hospital units classified into three groups or levels, depending on the sophistication and scope of the services they are equipped and staffed to provide. The distinctions among the three levels blur considerably in the field, however, and the resulting definitional confusions substantially complicate collecting and analyzing data.[3] Generally the results cited in this article are from nurseries that would qualify as Level III, the most sophisticated, although the level and criteria for classification often are not reported.

Remarkable improvements in infant mortality (deaths during the first year of life) have occurred in this country since 1915.[4-7] Most of this century's decline in infant mortality had taken place by 1950, however, and cannot be attributed to neonatal intensive care. The effects, if any, of newborn medical care only can be measured in terms of additional improvements in newborn survival subsequent to the introduction of modern methods. Between 1950 and 1965 the infant mortality rate continued to decline, but at a slower rate. During that period, nearly 60 per cent of the decline in infant mortality was experienced during the neonatal period (the first 28 days). This improvement in early survival is precisely the effect that would be expected if intensive newborn care was a major factor in improved survival. The end of this era coincides with the introduction of neonatal intensive care units (NICUs) and neonatology as a pediatric subspecialty.

Between 1965 and 1980 infant mortality fell by nearly 50 per cent, declining at 3.3 per cent annually, compared with the 1 per cent per year fall during the 1950–1965 period. Nearly all (88 per cent) of the decline since 1970 has occurred in the neonatal period, again consistent with a large medical care effect. Because the most recent improvements in the neonatal mortality rate were concurrent with the proliferation of intensive care newborn medical services, a number of studies have been undertaken to see whether the changes are, in fact, due to modern medical care.

The evidence collected falls into four categories: clinical trials, epidemiologic surveys, reports from individual intensive care units, and birthweight-adjusted mortality rate analyses. In reviewing this evidence, it is important to remember that each category addresses somewhat different aspects of intensive newborn care. Clinical trials of individual therapies are designed to test the efficacy of those measures in idealized situations. Clinical trials of a set of intensive measures compared with routine practice more closely would estimate the degree of effectiveness to be expected in general hospitals of the highest caliber. The reports from individual nurseries are a crude approximation of this same index of effectiveness. Epidemiologic surveys measure the net result on a community, including the practical issues related to access as well as the effectiveness of all available medical measures. By using an indicator that is related more directly to medical care than crude mortality rates, the birthweight-specific analysis provides a refinement on the epidemiologic surveys. Because the aspects of intensive newborn care assessed differ within each category they are presented separately.

Analysis of the Evidence on Effectiveness

Clinical Trials

There has been only one randomized controlled clinical trial of intensive care versus routine care of neonates.[8] That study, done in Australia in 1966–1970, showed improved survival for selected infants with very low birthweights, particularly those with respiratory distress syndrome (RDS), as experience in delivering intensive care developed. Kitchen and colleagues studied 238 of the 273 infants weighing between 1,000 and 1,500 grams who were born alive in the Royal Women's Hospital in Melbourne during that 4-year period. After excluding the 35 babies considered unsuitable for randomization, the authors assigned alternate patients to routine or intensive care. Intensive therapy consisted principally of umbilical artery cannulation for fluid and alkalai therapy and estimation of blood gases and acid-base status. Phototherapy, peripheral artery blood sampling, intermittent positive pressure ventilation (IPPV), and continuous positive airway pressure (CPAP) were not used for any infant.

Kitchen's results suggest that staff experience and newborn diagnosis are critical factors in determining outcomes. Improved survival was experienced by intensively treated patients throughout the study, but was statistically significant only during the latter 2 years. Even without the use of IPPV and CPAP, infants with RDS experienced significantly better survival with intensive care than with routine care in the latter period.

Had controlled trials been conducted in other nurseries during the late 1960s and early 1970s, there might be little question about the effect of intensive care methods on the survival of infants with very low birthweights. No other such studies have been identified, however. One leading neonatologist attributes this paucity of controlled studies to the notion that "controlled clinical trials of perinatal intensive care and regionalization have been neither feasible nor morally acceptable."[9] There have been clinical trials of some individual intensive care therapeutic techniques. One such study demonstrated improved mortality and shorter periods of hospitalization for infants who were transported under the care of a specialized neonatal team compared with those transported by the local ambulance authority.[10] The spectrum of specific intensive care measures studied in randomized clinical trials was reviewed recently by Sinclair and others.[11]

The debate over the efficacy of CPAP devices in the treatment of respiratory distress syndrome illustrates the status of most of the clinical evidence. This therapy has been hailed as a breakthrough, responsible for marked improvement in the survival of infants with RDS. However, the first controlled clinical trial was not conducted until several years after widespread adoption of CPAP and that study was not able to document statistically significant improvement in mortality.[12] Publication of that report in the pediatric literature was accompanied by a long editorial that reiterated the abundant evidence for the effectiveness of CPAP and detailed possible deficiencies of the controlled study, while expressing general support for clinical trials.[13] In short, many newborn therapies are unproven and even controversial. Thus no overall conclusions can be reached from the clinical literature alone.

Epidemiologic Surveys

Epidemiologic evidence suggests that medical care has helped lower mortality rates, but for the reasons discussed below the evidence is not conclusive. The Canadian experience showed a 50 per cent decline in neonatal mortality in Toronto in the decade following the organization of special neonatal care facilities.[14] An analysis of data for the United States between 1955 and 1973 demonstrated that, in general, mortality rates fell with increasing urbanization,[15] but urban areas were not stratified in terms of levels of intensive care available for newborns. Thus no conclusions could be drawn about the relationship between urban medical care and mortality.

In a major review article, Thompson and Reynolds summarized reports from various parts of the United States and Canada, all of which showed declines of 23 to 42 per cent after the development of regional intensive care facilities.[16] Although these data suggest that development of regional intensive care centers improved the outlook for infants in surrounding areas, the studies do not consider whether prematurity rates fell during the same period. Since declines in prematurity can markedly reduce neonatal mortality rates independent of changes in medical care, the omission greatly reduces the conclusions that can be drawn from these reports.

Sequential Reports From Individual Intensive Care Nurseries

Most of the literature on the effectiveness of neonatal intensive care consists of reports detailing the experience of individual nurseries. Twelve of these reports were reviewed by Thompson and Reynolds; all demonstrated reductions in inborn neonatal mortality, ranging from 17 to 76 per cent after the organization and designation of an intensive care newborn nursery.[16]

The impact that changes in prematurity rates can have on a nursery's mortality experience limits the value of these reports, however. If the distribution of births among birthweight groups changes from year to year, so that very small infants represent a lower proportion of total births, mortality experience will fall substantially. The proportion of infants in the very low birthweight category (1,500 grams or less) has the greatest impact on mortality rates.[17] Although those infants only represent about 1 per cent of all births, they account for approximately one half of all neonatal deaths.[3] Any individual nursery will have so few births in this category that even modest fluctuations in the proportion of infants with very low birthweights will have a marked impact on neonatal mortality.

The experience of the Bronx Municipal Hospital Center illustrates this phenomenon clearly.[18] In that facility, between 1966–1971 and 1972–1973, the in-hospital neonatal mortality rate declined from 16.9 per 1,000 live births to 11.2 per 1,000 live births. Over 75 per cent of that decline could be attributed to changes in the proportion of infants with very low birthweights (1,500 grams or less). That is, the neonatal mortality rate in the latter period would have been 12.6 per 1,000 live births just on the basis of a changed distribution of infants among birthweight categories.

The experience of the Bronx Municipal Hospital demonstrates that even in the absence of improvements due to medical care after birth, impressive declines in neonatal mortality can occur if a smaller proportion of high-risk infants is being born. Because the proportion of low birthweight infants (2,500 grams or less) declined by over 14 per cent between 1965 and 1978 in the United States,[19] reports of improvements in mortality must specify birthweight distribution if they are to help determine the effectiveness of newborn medical care. On the other hand, if mortality rates within specific birthweight groups do improve over time, this would be a strong indication that neonatal intensive care is having a beneficial impact. As Lee et al. have stated, if infants of 1,500 grams birthweight or less have a better chance of surviving after introduction of neonatal units, such a change must be due to "either improved neonatal medical care or to improved health status of the infants at birth, or both."[18]

A number of nurseries have reported mortality rates by birthweight before and after the introduction of an intensive care facility.[9, 20–26] The results for every nursery but one[26] show impressive declines in birthweight-specific mortality over time. Attributing most of these improvements to newborn medical care, Thompson and Reynolds felt that "NICUs have more than justified their reason for existence, reducing by approximately one-half the risk of neonatal death."[16]

Even birthweight-specific reports from individual nurseries are of limited usefulness, however, because of their small population samples. Often, only 20–30 infants in the very low birthweight groups are born during the study period, making it difficult to determine whether improvements over time are statistically significant. For this reason we grouped all available reports of mortality rates for inborn infants into 5-year periods beginning in 1961, recalculating published figures into comparable categories as necessary. These pooled data demonstrate significant declines over time for infants with very low birthweights (Table 1 & 2, Figure 1).Mortality has fallen from approximately 50 per cent to less than 20 per cent for infants with birthweights between 1,001 and 1,500 grams, while the extremely small newborns (1,000 grams or less) have improved from over 93 per cent mortality to just over 50 per cent.

These impressive trends over time demonstrated by the pooled birthweight-specific mortality data strongly suggest that

TABLE 1. Pooled Neonatal Mortality Data for Inborn Infants Weighing 1,000 Grams or Less

Hospital	1961–1965	1966–1970	1971–1975	1976–1980
University of California, San Francisco[22]	28/29	26/34	22/33	15/30
Children's, San Francisco[21]			30/35	8/24
University of Washington[29,30]		27/29		25/41
Medical Center, Columbus, Ga.[25]	157/168	27/34	26/35	
Bronx Municipal[18]		119/138	18/25	
University of Illinois[40]			29/38	
Harvard[41]				82/146
Cambridge Maternity[23]				9/25
Royal Victoria[9]		65/73	58/75	
Simpson Memorial, Edinburgh[42]		84/92		
University College Hospital, London[43]		29/36	26/33	
East Hospital, Göteborg, Sweden[24]		4/7		
Hammersmith Hospital, London[26]	37/41	40/46	20/23	
Total*	222/238 (933)	421/489 (861) †	229/297 (771) ‡	139/266 (523) §

Rates for each 5-year period differ significantly from those of each previous 5-year period.

* Numbers in parentheses are neonatal mortality rates (deaths/1,000 live births).

† $X^2c = 7.40$ ($p < 0.01$).

‡ $X^2c = 9.82$ ($p < 0.005$).

§ $X^2c = 37.18$ ($p < 0.001$).

intensive newborn medical care has helped to improve survival for infants with very low birthweights. All available series were included and it seems reasonable to assume that the range of evolving intensive therapies applied over each 5-year period was roughly comparable in the reporting nurseries. Thus, these pooled data reflect average results over time in intensive care nurseries. Nevertheless, these trends are based on nonexperimental, historic data and as such cannot be regarded as conclusive.

Comparison of the pooled data with the results reported by Kitchen et al. from Australia emphasizes the importance of clinical trials with contemporaneous controls.[8] Including the cases excluded from their study, Kitchen's group experienced a rate of 399 deaths per 1,000 live births for the entire population of infants in the 1,000–1,500 gram category born in their hospital. That result is comparable to the average rate of 372 deaths per 1,000 live births reported for intensive care by other nurseries during the same period, 1966–1970 (Table 2). Thus, during the years of a clinical trial, the experience of the Australian nursery as a whole was as good as that of other centers, even though it included infants who were

TABLE 2. Pooled Neonatal Mortality Data for Inborn Infants Weighing 1,001–1,500 Grams

Hospital	1961–1965 Deaths/Births	1966–1970 Deaths/Births	1971–1975 Deaths/Births	1976–1980 Deaths/Births
University of California, San Francisco[22]	26/55	23/63	15/80	9/68
Cambridge Maternity[23]				23/96
Royal Victoria[9]		25/89	14/90	
Harvard[41]				37/213
University of Washington[39]				14/75
Medical Center, Columbus, Ga.[25]	116/219	14/44	14/41	
Bronx Municipal[18]		69/193	11/42	
Simpson Memorial, Edinburgh[42]		71/154		
East Hospital, Göteborg, Sweden[24]		10/24		
Hammersmith Hospital, London[26]	38/85	40/81	34/81	
Royal Women's Hospital, Melbourne[44]	135/277	82/250		
Total*	315/636 (495)	334/898 (372)†	88/334 (263)‡	83/452 (184)§

Rates for each 5-year period differ significantly from those for each previous 5-year period.
* Numbers in parentheses are neonatal mortality rates (deaths/1,000 live births).
† $X^2c = 22.7$ ($p < 0.001$).
‡ $X^2c = 12.24$ ($p < 0.001$).
§ $X^2c = 6.73$ ($p < 0.01$).

allocated to routine care by the randomization process.

The controlled study design allowed Kitchen et al. to demonstrate the marginal value of applying specific additional intensive resources to infants with very low birthweights. During the last 2 years of that study, more than one third of the routinely treated group did not survive, compared with less than one fifth of the study infants. Thus, these figures allow a reasonably precise estimation of the value of the additional care, in contrast to the pooled nursery data. While the pooled data strongly suggest that the general improvements in survival over time represent the cumulative effect of new therapeutic modalities, they do not provide the direct evidence necessary to form such a conclusion with confidence.

Birthweight-Adjusted Mortality Rate Analyses

It is not possible to compare the experience in individual nurseries with the survival rates of infants of similar birthweight who were not cared for in intensive newborn units. As pointed out above, it is difficult to identify the levels of care available at different hospitals confidently. Moreover, not all states link birth certificates with death certificates, so it is impossible to know on a routine basis the birth-

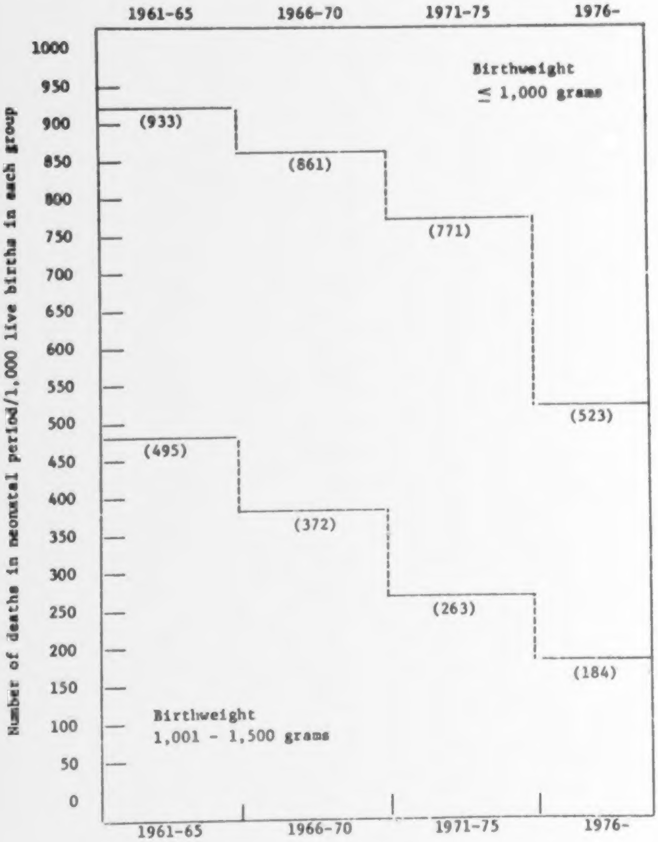


FIG. 1. Pooled neonatal mortality data.

weight of individual infants who die. Linked data for several areas have now been reported, however, and show major improvements in neonatal mortality between 1960 and the present.

Ronald L. Williams of the University of California, Santa Barbara, has analyzed changes in mortality rates among newborns in California over the past two decades.[27-30] Between 1960 and 1977 the neonatal mortality rate in California fell by 57 per cent. During the same period, low birthweight rates also declined by some 15 per cent. It was necessary, therefore, to partition the improvement in neonatal mortality into the part that was due to changes in birthweight distribution and the part due to other factors, including medical care. In contrast with the result noted above for Bronx Municipal Hospital Center, Williams showed that nearly 85 per cent of the remarkable decline in neonatal mortality in California resulted from birthweight-specific declines.[30] That is, the reduction in prematurity only accounted for about 15 per cent of the decline in neonatal mortality, while medical care and/or better health status of individual infants accounted for the remainder. More than half of the overall decline could be attributed to improvements in the mortality rates of infants with low birthweights. Williams also found that blacks experienced lower mortality rates than whites within low birthweight groups. This partially offset the fact that the birthweight distribution among blacks actually became less favorable over the 1960–1977 period.

Kleinman and others at the National Center for Health Statistics analyzed data from six states and reached similar conclusions.[31] They found that weight-specific changes accounted for nearly 80 per cent of the decline in early neonatal mortality rates. More than half of the overall decline resulted from improvements in mortality among the low birthweight groups. They estimated that, for the nation as a whole, about 83 per cent of the decline in early neonatal mortality for whites and 97 per cent of the decline for nonwhites could be attributed to factors other than changes in birthweight.

The same investigators discussed factors that could be causing the decline in birthweight-specific mortality rates. They rejected a suggestion that lower death rates were merely an artifact resulting from improved registration. They also concluded it was unlikely that any major shift in distinctions between live births and fetal deaths (stillbirths) had taken place, since fetal deaths were also declining. Finally, they attempted to deal with the question of whether infants within given birthweight groups might have different risk characteristics over time. They found that comparing gestational age distributions made no difference for the calculation and they concluded:

> . . . The most likely explanation for at least part of the recent decline in mortality levels among low birthweight infants is the development of the medical technology for the successful management of premature infants and the consequent proliferation of this technology. While there are many components of the technology which have not been critically evaluated, there are indications of effectiveness.[31]

Lee et al. examined the effect of birthweight changes on the improvements in neonatal mortality in the United States between 1950 and 1975. They determined that no improvement had taken place in the weight distribution or, with the exception of perinatal medical care, other factors known to affect survival.[32] Their conclusion, similar to those of Williams and Kleinman was that:

> [t]he most plausible explanation . . . is the steady improvement in perinatal medical care which has made for greater infant survival at a given birthweight.

Most recently, these same investigators used birthweight-adjusted analysis to examine variations in neonatal mortality among the 50 states and the District of Columbia, and among 13 industrialized nations.[17] They confirmed the role of the very low birthweight rate in predicting neonatal mortality; that rate accounted for about three quarters of the variance among the populations they studied. They suggested that variance of neonatal mortality after holding constant the very low birthweight rate might be a useful preliminary indicator of the quality of newborn care.

Discussion

Existing studies are helpful in answering part of the question regarding the role of medical care in bringing about improvements in infant mortality, but they do not settle all the issues. Of particular value are the birthweight-adjusted mortality rate analyses. By estimating that only 15 per cent of the decline in overall mortality rates is simply the result of changes in birthweight distribution, these analyses define the maximum impact on survival that could be attributed to medical care of the newborn. However, this is not to say that all of the remaining 85 per cent is necessarily a result of neonatal medical care.

One confounding problem would be any underlying changes in the health status of infants within each weight group that improved neonatal mortality independent of intensive newborn care. Such changes could be the result of medical or nonmedical factors. Medical factors, such as intensive intrapartum care or steroid therapy during premature labor to prevent hyaline membrane disease (HMD), could produce infants that are more likely to survive in the extremely fragile low birthweight groups. Obstetric care during pregnancy prior to labor would not be expected to affect birthweight-specific mortality rates but to reduce the chances of low birthweight.

Nonmedical factors such as improved maternal nutrition and socioeconomic status also have their primary beneficial effect through higher birthweights.[33,34] It is possible to speculate that better maternal nutrition and general health also result in subtle biochemical or other changes that further increase the chance of survival for infants in each weight group. Unfortunately, none of these hypotheses can be established or rejected confidently from the available data.

Another factor that complicates interpretation of the data is the learning curve associated with new technologies. The maximal impact of innovations may not appear for several years until staff expertise has developed. Similarly, favorable results in applying a new therapy to infants of a certain size may encourage its use in ever smaller infants, thus requiring additional time to see the maximal effect. Kitchen et al. noted that survival for routinely treated patients had improved over time even though all such infants were "managed in apparently the same fashion . . . "[8] They cautioned that this finding "illustrates the dangers inherent in drawing conclusions from retrospective or sequential data."

Similarly, one sketchy report describes a British hospital's experience in managing ill and premature babies with little or no intensive care.[35] No respirators, phototherapy, continuous positive airway pressure (CPAP) device, nor extensive mask and glove isolation were used in the nursery. Some modern techniques, including initial resuscitation, catheterization of umbilical vessels, exchange transfusions, and incubators were applied. The absolute levels of mortality achieved in each weight category would have been unacceptable at a major nursery in the United States during the same time period; however, there was significant annual improvement, even within birthweight groups. On the basis of this limited report, one is not able to conclude whether this nursery is, in fact, able to improve mortality without changing neonatal care and without delivering the most advanced methods. Such reports do serve to illustrate the continuing need for new evidence to determine the precise contribution of neonatal care in improving birthweight-specific mortality rates.

The work by Williams provides some further indirect evidence that supports the hypothesis that medical care was a major factor in lowering mortality rates. For infants in all hospitals in California, he found that once the risks faced at each hospital are taken into account, hospitals with neonatal intensive care nurseries have better survival rates than those without intensive care units, and hospitals with large delivery services and high specialist–generalist ratios have the best performances.[28,36]

In summary, while it may not be possible to determine the precise role that neonatal intensive care has played in lowering mortality rates among low birthweight infants, the evidence supports the conclusion that neonatal care has been a significant factor. The degree, if any, to which infants are somehow improving their inherent ability to survive from year to year and the role of intensive intrapartum obstetrical care could be tested only through additional controlled clinical trials.

Conclusion

Present knowledge provides only tentative conclusions on the effectiveness of neonatal intensive care. Nationally, infant mortality has fallen by an impressive 49 per cent (from 24.7 per 1,000 live births in 1965 to 12.5 per 1,000 live births in 1980) during the recent period when access to increasingly sophisticated care grew substantially. Among the factors that may have led to this dramatic decline are improved prenatal and maternity care, lower fertility rates and improved maternal nutrition, as well as improvements in the medical care provided to ill or extremely small infants. It has been difficult to assess the impact of medical care because of inadequate information on the actual types of neonatal intensive care provided and their relationship to medical outcome.

Within these limitations, partial answers to the questions of the value of medical care for newborns are available at present. These answers suggest that medical care of the newborn has been effective in many ways, but present information is inadequate to determine the degree to which the amount or type of medical care is associated with improved outcome. Much of the improvement in mortality rates for newborns in the United States has been in the areas where medical care would be expected to have its greatest impact: survival of infants with very low birthweights (less than or equal to 1,500 grams, or 3.3 pounds) has risen from about one fourth to over two thirds of infants of that size, and virtually all of the improvement in survival occurred during the first 28 days of life. The solitary clinical trial supports the crude epidemiologic data, the individual nursery success stories, and the encouraging birthweight-adjusted time series data. Together, this constitutes a body of highly suggestive evidence that intensive care of the newborn has had beneficial effects on survival.

The results of previous studies, then, tell us that more intensive care generally is associated with better survival, but they do not help us determine the critical elements of such care that make a difference. Existing studies treat intensity of medical care (defined as type and amount of services provided) as a "black box," failing to distinguish the roles of case-mix, staffing, treatment modes, size, utilization, and the other variables that comprise intensity. The present system of categorizing nurseries into different levels of intensity is not precise enough nor sufficiently standardized, so that great variation exists among nurseries. Consequently, the levels of intensity and the possible relationship between the intensity of care and outcomes cannot be specified in ways that would be useful for planning or making decisions about regionalization or about relating reimbursement to levels of care or other indicators of cost.

Resolution of the remaining questions will require a multi-tiered strategy. Studies such as the Australian controlled trial would violate current standards of care and probably can no longer be attempted for ethical and legal reasons. Individual unproven technologies, as well as future innovations, could and should be evaluated through prospective controlled clinical trials. Retrospective and current analyses testing the relationship between survival rates, and the diffusion of advanced personnel and equipment would be a valuable contribution. In particular, studies that analyze the relationship between specific newborn intensive care resources and health outcomes for selected conditions would provide new information on the impacts of specific organizational and technologic components of medical care. The high personal and financial costs of providing intensive newborn care require more definitive evidence concerning the effects of such care.

Finally, an equally important question left unanswered by survival figures is whether the increased number of infants being saved will be able to lead normal lives, or whether they will suffer from serious handicaps such as cerebral palsy and mental retardation. Other serious health problems of survivors may include high rates of child abuse; learning, language and behavior disorders; and complications of therapy such as blindness due to retrolental fibroplasia and impaired hearing secondary to the use of certain antibiotics. These questions are perhaps even more perplexing than the issue of survival because the absolute number of afflicted individuals can increase at the same time that the incidence of individual problems is decreasing. Such a paradoxical result will occur if mortality rates fall faster than the rates of decline in the incidence of residual disorders. This aspect of intensive care is being addressed,[37,38] but is still in need of additional long range, intensive analysis.[3]

Acknowledgments

Special thanks are due to Nancy Barrand, LuAnn Heinen, and Suzanne Stenmark for their contributions, to Carol Bok, Barbara Johnson, and Les Gates for their preparation of the manuscript, and to our colleagues at the Institute for Health Policy Studies, particularly Steven A. Schroeder, for reviewing the earlier drafts.

References

1. Silverman WA. Incubator-baby sideshows. Pediatrics 1979;64:127.

2. Schlesinger ER. Neonatal intensive care: planning for services and outcomes following care. J Pediatr 1973;82:916.

3. Budetti P, McManus P, Barrand N, Heinen L. The costs and effectiveness of neonatal intensive care. Washington, D.C.: Congressional Office of Technology Assessment. U.S. Government Printing Office, 1981.

4. National Center for Health Statistics. Vital statistics of the United States 1975, vol 2 Mortality, part A. Hyattsville, Md.: National Center for Health Statistics, 1979. (DHEW Publication No. (PHS) 79-1114).

5. National Center for Health Statistics. Final mortality statistics 1977. Hyattsville, Md.: National Center for Health Statistics, 1979. (Monthly Vital Statistics Report, Advance Report. DHEW Publication No. (PHS) 79-1120, vol 28, No 1 Suppl.).

6. National Center for Health Statistics. Final mortality statistics 1978. Hyattsville, MD: National Center for Health Statistics. 1980. (Monthly Vital Statistics Report, Advance Report. DHHS Publication No. (PHS) 80-1120, vol 29 no 6 Suppl. 2).

7. National Center for Health Statistics. Births, marriages, divorces and deaths for 1980. Hyattsville, Md.: National Center for Health Statistics, 1980. (Monthly Vital Statistics Reports, Provisional Data. DHHS Publication No. (PHS) 81-1120, vol 29, no 12).

8. Kitchen WH, Ryan MM, Rickards A, et al. A longitudinal study of very low-birthweight infants. I. Study design and mortality rates. Develop Med Child Neurol 1978;20:065.

9. Usher R. Changing mortality rates with perinatal intensive care and regionalization. Seminars Perinatol 1977:1:309.

10. Chance GW, Matthew JD, Gash J, Williams G, Cunningham K. Neonatal transport: a controlled study of skilled assistance. J Pediatr 1978;93:662.

11. Sinclair JC, Torrance GW, Boyle MH, Horwood SP, Saigal S, Sackett DL. Evaluation of neonatal-intensive-care programs. N Engl J Med 1981;305:489.

12. Belenky DA, Orr RJ, Woodrum DE, Hodson A. Is continuous transpulmonary pressure better than conventional respiratory management of hyaline membrane disease? A controlled study. Pediatrics 1976;58:800.

13. Wung JT, Stark RI, Hegyi T, Driscoll J, James LS. CDP: a major breakthrough! Pediatrics 1976;58:783.

14. Swyer PR, Goodwin JW, eds. Regional services in reproductive medicine. Toronto: The Joint Committee of the Society of Obstetricians and Gynecologists of Canada and the Canadian Pediatric Society, 1973.

15. Eisner V, Pratt NW, Hexter A, Chabot MJ, Sayal N. Improvement in infant and perinatal mortality in the United States, 1965–1973: I. Priorities for intervention. Am J Public Health 1978;68:359.

16. Thompson T, Reynolds J. The results of intensive care therapy for neonates. Perinatal Med 1977:5:59.

17. Lee KS, Paneth N, Gartner LM, Pearlman M. The very low birthweight rate: principal predictor of neonatal mortality in industrialized populations. J Pediatr 1980;97:759.

18. Lee KS, Tseng PI, Eidelman AI, Kandall SR, Gartner LM. Determinants of the neonatal mortality. Am J Dis Child 1976;130:842.

19. Bureau of the Census. Statistical abstract of the United States 1980. Washington, D.C.: U.S. Government Printing Office, 1980. (Table 91.)

20. Alden ER, Mandelkorn T, Woodrum DE, Wennberg R, Parks C, Hodson WA. Morbidity and mortality of infants weighing less than 1,000 grams in an intensive care nursery. Pediatrics 1972;50:40.

21. Hirata T, Halliday HL. Survival of infants of 1,000 g birth weight or less. (Letter), Br Med J 1977;2:1670.

22. Phibbs RH. Survival rates by birthweight. San Francisco: University of California Nurseries, 1979. (Unpublished).

23. Robertson NRC. Letter. Lancet 1979;ii:362.

24. Sabel KG, Olegard R, Victorin L. Remaining sequelae with modern perinatal care. Pediatrics 1976;57:652.

25. Thompson RE, Souma ML, Cassady G, Sumners J. Impact of intensive newborn care on neonatal mortality in a community hospital. South Med J 1976;69:688.

26. Jones RA, Cummins M, Davies PA. Infants of very low birthweight: a 15-year analysis. Lancet 1979;i:1332.

27. Williams RL. Intrauterine growth and neonatal risk in California. (Appendix) Cunningham GC, Hawes W, Madore C, eds. Santa Barbara: University of California Community and Organization Research Institute, 1976.

28. Williams RL. Measuring the effectiveness of perinatal medical care. Med Care 1979;17:95.

29. Williams RL, Cunningham GC, Norris FD, Tashiro M. Monitoring perinatal morality rates: California 1970 to 1976. Am J Obstet Gynecol 1980;136:559.

30. Williams RL, Chen P. Identifying the sources of the recent decline in perinatal mortality rates in California. N Engl J Med 1982;306:207.

31. Kleinman JC, Kovar MG, Feldman JJ, Young CA. A comparison of 1960 and 1973–1974 early neonatal mortality in selected states. Am J Epidemiol 1978;108:454.

32. Lee KS, Paneth N, Gartner LM, Pearlman M. Neonatal mortality: an analysis of the recent improvement in the United States. Am J Public Health 1980;70:15.

33. National Center for Health Statistics. Teenage childbearing: United States, 1966–75. Hyattsville, Md.: National Center for Health Statistics, 1977. (Monthly Vital Statistics Report, Natality Statistics. DHEW Publication No. (PHS) 77-1120, vol 26, no 5 Suppl.).

34. Taffel S. Factors associated with low birthweight, United States, 1976. Hyattsville, MD: National Center for Health Statistics, 1980. (Series 21, No. 37).

35. Hughes-Davies TH. Conservative care of the newborn baby. Arch Dis Child 1979;54:59.

36. Williams RL. Measuring the effectiveness of perinatal medical care. Santa Barbara: University of California Community and Organization Research Unit, 1980. (U.S. Bureau of Community Health Services Grant Report MC-R-060390-03).

37. Shapiro S, McCormick M, Starfield B, Krischer J, Bross D. Relevance of correlates of infant deaths for significant morbidity at 1 year of age. Am J Obstetr Gynecol 1980;136:363.

38. McCormick M, Shapiro S, Starfield BH. Rehospitaliziation in the first year of life for high-risk survivors. Pediatrics 1980;66:991.

39. Jackson NS. Unpublished data. Division of Neonatal Biology. Department of Pediatrics, University of Washington, 1979.

40. Bhat R, Raju TNK, Vidyasagar D. immediate and long-term outcome of infants less than 1,000 grams. Crit Care Med 1978;6:147.

41. Brown ER, Taeusch HW. Letter. Lancet 1979;ii:362.

42. Drillien, CM. Aetiology and outcome in low-birthweight infants. Dev Med Child Neurol 1972;14:563.

43. Stewart AL, Turcan D, Rawlings G, Reynolds EOR. Prognosis for infants weighing 1,000 g or less at birth. Arch Dis Child 1977;52:97.

44. Kitchen W, Campbell D. Controlled trial of intensive care for very low birthweight infants. Pediatrics 1971;48:711.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Improving Physician Compliance With Preventive Medicine Guidelines

DAVID I. COHEN, M.D., M.SC.,* BENJAMIN LITTENBERG, B.A.,† CHERYL WETZEL, B.A.,‡ AND DUNCAN vB. NEUHAUSER, PH.D.§

Similar general medical outpatient clinics with randomly assigned patients were used to evaluate the effectiveness of a program that was to increase house staff compliance with preventive medicine guidelines. Two clinics were designated experimental and two served as controls. In the experimental clinics, age-specific checklists of all recommended preventive procedures (drawn from the Canadian Task Force report on The Periodic Health Examination and American Cancer Society guidelines) were appended to each patient's chart. In addition, house officers were presented with a series of weekly seminars dealing with issues in screening, as well as the specific recommendations included in the checklist. House officers in all four clinics were tested for their knowledge and attitudes toward the preventive program before and after the intervention. Counts of immunizations and mammograms performed and the total populations eligible for these procedures were determined for all four clinics. As predicted, test scores as well as mammography and immunization rates increased significantly (from 2–10 per cent) in the intervention clinics as compared with controls. We conclude that this intervention was clearly effective in the short run. However, follow-up studies will be necessary to determine whether the desired long-term effect has been achieved.

TOO LITTLE PREVENTIVE care is provided. What can be done to change the behavior of physicians responsible for providing such care? The academic medical center has been recognized as a hub for diffusing information and technology. It is an environment that fosters the acquisition of knowledge and the shaping of attitudes. A number of studies have examined the impact of various maneuvers upon house officers' use of laboratory testing. Several of these studies are summarized in the first of our "Metro Firm Trials" published in this journal.[1] Most studies of physician behavior change have explored the apparent overutilization of diagnostic testing in the teaching hospital setting. We, however, focused our attention upon the use of screening and preventive services, which have been underutilized by all indications.[2] A recent review by Carter, Belcher, and Inui[3] has chronicled some of the explanations for this underutilization, which include physician, patient, societal, and economic factors.

In this study, we test under controlled conditions and show the effectiveness of a simple intervention to improve house of-

---

* Assistant Professor, Department of Medicine, Case Western Reserve University, Chief, Division of General Medicine, Cleveland Metropolitan General Hospital, Cleveland, Ohio.

† Medical student, Case Western Reserve University, Cleveland, Ohio.

‡ Research assistant, Cleveland Metropolitan General Hospital, Cleveland, Ohio.

§ Professor, Department of Community Medicine and Epidemiology, Case Western Reserve University, Cleveland, Ohio.

Supported by Cleveland Foundation and Primary Care Funds, State of Ohio Board of Regents.

Portions of this article were presented at the Midwestern Regional Meeting of the American Federation of Clinical Research and at the National Meeting of The Robert Wood Johnson Foundation Clinical Scholars Program.

Reprint requests: David I. Cohen, M.D., M.Sc., Cleveland Metropolitan General Hospital, 3395 Scranton Road, Room H-331, Cleveland, OH 44109.

0025-7079/82/1000/1040/$01.55 © J. B. Lippincott Co.

ficer compliance with preventive medicine guidelines. We observe the relationship between physicians' knowledge and attitudes, and their use of preventive testing.

The Setting

The study was conducted in the Medical Outpatient Department at Cleveland Metropolitan General Hospital in the fall, 1980.

The Department of Medicine was reorganized during 1975 into a system composed of four general medical firms,[4] each consisting of a 28-bed inpatient unit as well as its own outpatient clinic. House officers are randomly assigned to a firm during internship and remain with that firm for the duration of their training. Patients similarly are randomized into one of the four firms, either on inpatient admission or on first contact with the outpatient area. They are assigned to a specific house officer with whom they remain associated until the house officer leaves the program. Then they are reassigned to another house officer in the same firm. The system thus has provided us with a measure of continuity not generally available in most academic medical centers.[5] Continuity is seen in the inpatient and outpatient services and has enhanced physician–patient relationships, clinic and ward team associations, and faculty–house staff teaching relationships. Moreover, the firm system has provided us with a unique laboratory for clinical and health care research. In essence what we have are four similar groups of patients and physicians, enabling us to study the effects of a maneuver in one or more films, while using the others as controls.

Methods

For this study two firms were randomly designated experimental and one served as a control. Three age-specific checklists, each delineating recommended preventive screening tests and procedures for all patients falling within a given age group, were designed. Items included in the lists were limited to those given Class A recommendations by the Canadian Task Force on the Periodic Health Examination[6] and those recommended by the American Cancer Society (Table 1).[7]

On the experimental firms the appropriate checklist (by age) was affixed to the chart cover of each patient who came to the medical clinic to serve as a reminder of the appropriate preventive measures for that patient. A trained research assistant quickly scanned each medical record to determine whether any of the recommendations had been followed in the recent past. Her findings were recorded on the checklists and appropriate radiology requests were completed as indicated. The house officer who was responsible for the

TABLE 1. Items Included on Preventive Medicine Checklists

	Age 16–44	Age 45–64	Age 65+
Required	Check blood pressure Pap test Breast exam Tetanus immunization	Check blood pressure Pap test Breast exam Mammogram (50–60) Stool hematest Tetanus immunization	Check blood pressure Stool hematest Pneumovax Influenza immunization Tetanus immunization

The following are optional for high risk groups at physicians' discretion: PPD, VDRL, GC culture.

primary care of the patient then was left to determine which procedures should be followed. Final orders for all screening tests and immunizations thus were made by the primary care physician.

The checklist intervention proceeded for a period of 4 months. Specific counts of the vaccinations and mammograms performed were kept by the clinic nursing staff and the radiology department for the two experimental firms and one control firm. The numbers of patients seen in the clinics who were eligible for these procedures were compiled by a research assistant. The percentages of eligible patients who actually received mammograms, influenza immunizations, and pneumovax during the period of the study were tabulated with these data. Furthermore, these data enabled us to identify previously screened patients to prevent duplications.

At the time the checklists were introduced into the medical clinics, a series of five seminars dealing with issues of screening and preventive medicine was offered to the medical house staff. These seminars were given weekly during the first 2 months of the clinic intervention. The first session was devoted to a general discussion of screening and of the methods employed to determine the efficacy of various screening maneuvers and preventive interventions. Subsequent sessions dealt with the specific recommendations being encouraged in the medical clinics. While house officers from all firms were invited to these sessions, those from the experimental firms were particularly encouraged to attend. In fact, with only one exception, no house officers from the control firms attended, and even attendance by members of the experimental firms was poor, averaging 1.2 seminars per physician for the 22 physicians in these firms.

A test instrument was developed to assess physicians' knowledge of and attitudes toward the use of preventive procedures and screening maneuvers. This test was administered to all medical house officers and faculty prior to the experimental intervention and seminar series, and again to all house officers at the conclusion of the study. The test consisted of two parts, one dealing with factual information drawn directly from the Report of the Canadian Task Force on the Periodic Health Examination and from the American Cancer Society guidelines, and the other dealing with attitudinal issues concerning physicians' choices of screening tests in varied clinical situations. Positive scores for the attitudinal section were given for those screening tests judged appropriate for a given situation by a panel of experts; negative scores were given for those judged inappropriate.

Results

The results of the intervention are summarized in Table 2. Over the 4-month period of the study only 5 per cent of the eligible controls received pneumovax, while 42 per cent of patients on the experimental firms were immunized. Influenza immunizations were delivered to 4 per cent of the eligible population on the control firm and 36 per cent of the experimental group. Similarly, mammography was performed in only 4 per cent of the control group and 32 per cent of eligible patients in the experimental firms. All differences are significant at the $p < .001$ level.

The data are graphically presented in terms of monthly rates in Figure 1. Virtually none of our patients routinely received screening mammograms or immunizations at the start of our intervention. This occurred partly because the influenza vaccine for that year was released just at the beginning of the study. The effects of the intervention, however, were seen immediately. Pneumovax and influenza immunization rates rose 30 per cent on the intervention firms during the first month of the study and continued to rise to almost 60 per cent during the subsequent 2 months,

TABLE 2. Immunization and Mammography Rates Among Eligible Patients During 4-Month Intervention

	Experimental Firms (8B & 11C)		Control Firms (6C)	
	Number Eligible	% Completed	Number Eligible	% Completed
Pneumovax	547	42%	291	5%
Influenza	581	36%	291	4%
Mammogram	290	32%	138	4%

$p < 0.001$ in all cases.

finally dropping during the fourth month. Rates on the control firm remained less than 10 per cent for the duration of the study. Similar results were seen in the case of mammography screening. It should be noted that it was impossible to eliminate previously screened patients from our tabulation of denominators, that is, the populations eligible for immunizations and mammograms. Therefore rates reflect the most conservative possible estimates of completed procedures. Furthermore, because patients tend to return to our medical clinics at about 3-month intervals, it is conceivable that the drop noted toward the end of the study represents the return of patients who had already undergone appropriate screening and vaccination for that year.

The data presented up to this point have been analyzed in terms of the patient population on the firms. Because our intervention was directed toward improving the delivery of preventive measures to our patients, we felt that the patient population was an appropriate unit of analysis to evaluate its effectiveness. However, our interest in the effect of our intervention upon physicians' attitudes and behaviors necessitated that the data also be examined using the physician as unit of analysis.

Differences between mean scores for delivery of mammography, influenza immunization, and pneumovax by physicians on the experimental and control firms all were significant at the $p < .001$ level. Similarly, point-to-serial correlations using Pearson's R showed highly significant differences between experimental and control groups in support of the experimental hypotheses. Membership on the experimental firms was correlated strongly with the delivery of mammography ($r = 0.53$, p

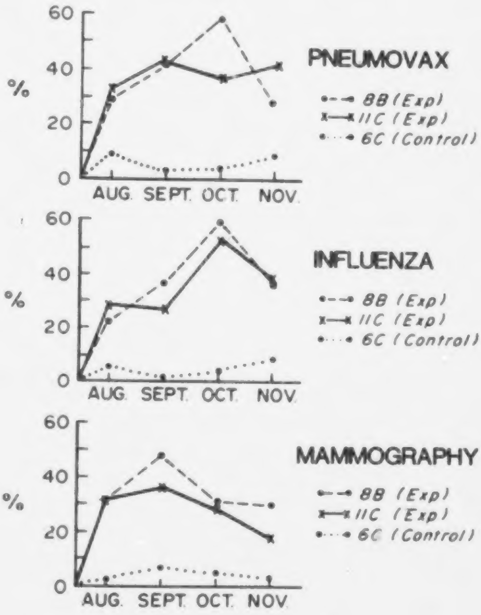


FIG. 1. Percentage of eligible patients receiving immunizations and mammograms: comparisons of experimental and control firms.

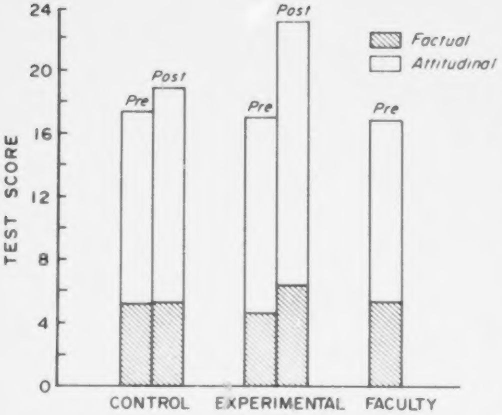


FIG. 2. Comparison of pretest and posttest scores between experimental and control firms.

< 0.001), influenza immunizations ($r = 0.65$, $p < 0.0001$), and pneumovax ($r = 0.70$, $p < 0.0001$).

Scores of the pretest and posttest were tabulated following the clinic intervention to assure that there was no bias in terms of faculty–house staff interactions. Three separate scores were calculated, one for the total test, a perfect score being 36; one for the attitudinal portion alone, a perfect score being 27; and one for the factual portion alone, a perfect score being 9. The results are summarized in Figure 2. No statistically significant differences were observed between the mean factual, attitudinal, or total scores of control and experimental groups on pretesting. Similarly, no differences were noted between house staff and faculty mean scores on pretesting. However, a significant difference was observed in the mean attitudinal and total test scores ($p < 0.05$ in both cases) on posttesting between the experimental and control groups. Similarly, the differences between experimental and control groups in the improvement in total scores from pretest to posttest was significant ($p < 0.05$). Unexpectedly, no difference was observed in the mean posttest factual scores between experimental and control groups.

An examination of point-to-serial correlations revealed that among physicians in experimental firms, improvement in scores from pretest to posttest was strongly related to seminar attendance ($r = 0.67$, $p < 0.0006$) but was not correlated with the use of preventive procedures. Seminar attendance alone similarly was not correlated with the use of preventive procedures.

Correlations between the use of the three preventive measures were high (Mammography and influenza vaccine $+0.49$ $p = 0.002$; mammography and pneumovax $+0.52$ $p = 0.001$; and pneumovax and influenza $+0.91$ $p < 0.0001$).

Discussion

Much of current medical practice is based upon precedent alone. The application of scientific rigor to the evaluation of the effectiveness and efficacy of various preventive and therapeutic regimens, as well as of systems for their implementation is a relatively recent development. Not surprisingly, in this context evaluative data often has had less clinical impact than may be appropriate. Consequently we frequently observe high utilization of interventions of questionable efficacy, while those of demonstrated value may be consistently underutilized. The causes of this discrepancy between ideal and actual practice are manifold and relate to issues of technology diffusion, knowledge, and attitudes of practitioners and patients, and to a wide variety of incentives determined by the personal, professional, social, and economic environment.

In this study we examined the use of a small number of preventive interventions of well-documented efficacy.[6–8] These procedures and preventive maneuvers have been evaluated carefully in the literature and were found to be safe and effective. Guidelines for their use have been clearly defined and target populations upon whom specific interventions would have the most impact have been determined.

Because the effectiveness of these preventive interventions has been well documented in the literature, we felt that it was safe to assume that the process of immunization and screening would result in desired outcomes. The efficacy of the recommended screening procedures has been demonstrated in detecting diseases at a point early enough to alter their natural histories. Few medical practices have been submitted to such intense scrutiny. It follows that compliance with these guidelines should be part of standard medical practice. Surprisingly, this has not been the case. Confusion exists with respect to what items should be included in the periodic health examination and the frequency with which they should be performed.[2]

In our clinics we found that, with the exception of blood pressure measurements, routine screening and preventive procedures rarely were carried out. We therefore proceeded to design an intervention to improve house staff compliance with preventive medicine guidelines. The intervention itself purposely was kept simple, relying upon a minimum of on-site staff supervision and monitoring. What we observed was a rather remarkable change in the house officers' attitudes and use of preventive procedures. It was interesting to see that the effect of the educational seminars seemed far less important than the simple checklists in changing physician behaviors. It appeared that a change in attitude, rather than in specific factual knowledge, was correlated most closely with the change in use of preventive procedures. Further research will be necessary to determine what factors influence attitude most directly. Similarly, further investigation will be necessary to determine the duration of the effect of our intervention.

It may be, however, that an intervention of the sort we have described should be repeated on an annual basis in teaching hospitals. As with all new information and diffusing technologies, simple knowledge of efficacy and effectiveness does not ensure appropriate use. A constant reminder, perhaps in the form of a checklist, may be necessary to improve the delivery of certain procedures that should be, but often are not, regularly considered by physicians. A system to ensure the periodic application of screening and preventive maneuvers of demonstrated effectiveness clearly can contribute to the provision of quality care.

References

1. Cohen DI, Jones P, Littenberg B, Neuhauser D. Does cost information availability reduce physician test usage? A randomized clinical trial with unexpected findings. Med Care 1982;20:206.

2. Romm FJ, Fletcher SW, Hulka BS. The periodic health examination: Comparison of recommendations and internists' performance. South Med J 1981;74(3):265.

3. Carter WB, Belcher DW, Inui TS. Implementing preventive care in clinical practice: problems for managers, clinicians, and patients. Med Care Review 1981;38(4):195.

4. Waggoner DM, Frengley JD, Griggs RC, Rammelkamp CH. A "firm" system for graduate training in internal medicine. J Med Educ 1979;54:556.

5. Lampman JH. Unpublished data.

6. Spitzer WO. Report of the Task Force on the Periodic Health Examination. Can Med Assoc J 1979;121:1193.

7. American Cancer Society. Guidelines for the cancer related checkup. Recommendations and rationale. CA-A Cancer Journal for Clinicians 1980;30(4):194.

8. Breslow L, Somers AR. The lifetime health-monitoring program. N Engl J Med 1977;296:601.

MEDICAL CARE
October 1982, Vol. XX, No. 10

Communication

Spending Time With Patients

The Impact of Organizational Structure on Medical Practice

FREDRIC D. WOLINSKY, PH.D.,* AND WILLIAM D. MARDER, A.M.†

This article assesses whether the amount of time that physicians spend overall, in the office, and in the hospital per patient visit depends upon the organizational structure of the physician's practice (i.e., does the physician practice in the fee-for-service system or in an HMO, and if in an HMO, in what type of an HMO). Data pooled from two national studies (N = 2,521) reveal several interesting patterns, including: 1) on all measures, solo physicians spend more time per patient visit than physicians in group practice; 2) overall and in the office, internists spend more time per patient visit than general practitioners; 3) overall and in the hospital, obstetricians-gynecologists spend more time per patient visit than general practitioners; 4) overall and in the office, physicians in prepaid group practices spend less time per patient visit than physicians in fee-for-service group practices; 5) overall and in the office, physicians reimbursed on a straight salary basis spend more time per patient visit than nonsalaried physicians; and, 6) in the hospital, physicians in staff model HMOs spend more time per patient visit than physicians in fee-for-service group practice. The implications of these findings for future studies of the effects of organizational structure on medical practice are discussed.

GIVEN THE EMPHASIS on improving access to ambulatory medical care while containing the rapidly rising costs of such care, Health Maintenance Organizations (HMOs) are receiving more currency as the most palatable competitive approach for the 1980s.[1,2] Indeed, the health services research literature is replete with current and extensive assessments of the promise and performance of HMOs.[3-6] One of the most commonly assumed yet seldom documented attributes of HMOs is the fact that HMO enrollees must wait longer to see a physician and then get to spend less time with the physician than do their fee-for-service counterparts.

In one of the rare comparative studies of this alleged attribute, Mechanic found that "fee-for-service physicians spend more time in direct patient care activities than those in prepaid practice, and devote more time to each patient."[7,8] The explanation

---

* Associate Professor, Hospital and Health Care Administration; Director, Doctoral Program in Health Services Research, Saint Louis University Medical Center.

† Research Associate, Center for Health Policy Research, American Medical Association.

The views presented in this paper are those of the authors and do not necessarily represent the official policy or position of either the Saint Louis University Medical Center or the American Medical Association.

Reprint requests: Fredric D. Wolinsky, Ph.D., Saint Louis University Medical Center, 3525 Caroline Street, St. Louis, Missouri 63104.

0025-7079/82/1000/1051/$01.25 © J. B. Lippincott Co.

he gives for this generally accepted finding is that[8]:

> . . . the patient load characteristic of general practice in pre-paid groups encourages a more assembly line practice which is less responsive to patients than the pattern characteristics of fee-for-service practice. Pre-paid physicians work during scheduled hours and may deal with increased load by processing patients more rapidly. Fee-for-service physicians tend to respond to increased demand by working longer hours.

Two tacit assumptions are operative here. On the one hand, it is assumed that physicians in prepaid group practices have no financial incentive to adjust to increased patient demand by lengthening their work week. On the other hand, it is assumed that fee-for-service physicians can adjust to increased patient demand by lengthening their work week because they will receive additional compensation.

Essentially, this argument assumes the existence of what may be called a target-time hypothesis. That is, although physicians are willing, for the most part, to adjust their practice schedules to patient demands, they will do so only within what they consider to be a reasonable range of hours worked per week. Based on this tacit assumption and in the absence of their receiving additional compensation for practicing additional hours beyond the reasonable range, Mechanic posited that: 1) salaried physicians would adjust their practices to the increase in patient demand by reducing the amount of time that they spend per patient, while 2) physicians in fee-for-service practice would respond by lengthening their work week in return for additional compensation. The data that Mechanic presented were consistent with his expectations,[7,8] although seemingly comparable unpublished data from Held and Reinhardt's Group Practice Study indicate that office visits are generally "about the same length in both practice modes."[9]

In a comparative study of the medical practices of physicians in different types of HMOs,‡ Wolinsky and Corry examined variations in the amount of time these physicians spent with patients.[10] Their findings indicate that in staff model HMOs (wherein physicians are likely to be on a straight salary) physicians spent more time per patient in the office and in the hospital than physicians in either group model or independent practice associations (IPA) model HMOs. Specifically their data indicate that the staff model HMO physicians spent from 10 to 20 per cent more time with their patients than did other HMO physicians.

---

‡ Basically, there are three types of HMOs: staff models, group models, and individual practice associations or IPAs. In a group model HMO all of the physicians are members of the same group practice. Their medical group contracts with the HMO to provide its services, and is thus a separate legal and fiscal entity from the HMO. Usually, the group services that HMO exclusively, and either owns or contracts with the hospital(s) to which its patients are admitted. Members of the group are reimbursed on a salary or salary plus profit-sharing basis. The key element, then, in group model HMOs is that the physicians have a clear proprietary interest in the success or failure of the group and of the HMO. In essence, they are partners in its financial success or failure. The classic example of group model HMOs is the Kaiser Foundation Health Plan.

Unlike physicians in group models, physicians in staff model HMOs are employed directly by the HMO itself. As a result they are employees rather than partners. Therefore, they do not have as clear a proprietary interest in the HMO. They are usually reimbursed on a salary basis, although adjustments may be made based on average patient loads. The classic example of the staff model HMO is the Group Health Plan of Puget Sound.

The IPA is a loose federation of independent, individual physicians who agree to treat patients enrolled in a third party's HMO on a fee-for-service basis in their own offices. Frequently, the IPA is jointly sponsored by the local medical societies and insurance companies. When an IPA physician provides services to an HMO enrollee, the physician is reimbursed by the IPA for services rendered, less a modest discount rate for the administration of the plan. Unlike staff and group model HMO physicians, who usually have the majority of their patients coming from the HMO, IPA physicians usually have no more than 10 per cent of their patients coming from the HMO. As a result, while the HMO practice is usually the major if not the entire practice for staff and group model physicians, it is a small sideline for physicians in IPAs. The classic example of an IPA is the Forbes Health Maintenance Plan of Pittsburgh.

In explaining their findings, Wolinsky and Corry posited that the longer patient visits provided by staff model HMO physicians reflect the absence of financial incentives motivating them to provide either more services in general or fewer services per enrollee, while seeing more enrollees. Accordingly, they argue that:

> . . . because staff model HMO physicians are principally reimbursed on a straight salary, there is no financial incentive impinging upon them to reduce the amount of time they spend with each patient. As a result, staff model HMO physicians appear to be more willing to spend significantly longer periods seeing their patients in both the office and hospital settings.[10]

Thus, like Mechanic, Wolinsky and Corry focus on an economic explanation, although their emphasis is on the absence rather than the presence of financial incentives that would motivate staff model HMO physicians to increase the demand for their services.

Clearly the data presented by Mechanic and by Wolinsky and Corry contradict each other and lead to different, albeit economically oriented, interpretations. Indeed, both interpretations appear to be straightforward and plausible, but a theoretic reconciliation is not obvious. Empirically, however, the results of the two studies may vary because of differences in design that lead to alternative analytic shortcomings. For example, Mechanic's study directly compares fee-for-service physicians with physicians in prepaid group practices, assuming that all HMOs are basically alike and modeled after the traditional group model typified by Kaiser Permanente. Wolinsky has argued that this is not the case and that the differences in structural configuration between staff model, group model, and IPA model HMOs have significant impacts on both the macro and micro level performance of the plan and its individual physician's practices.[6] The Wolinsky and Corry study supports this claim by showing significant differences in individual physician's medical practices between HMO types. Their data, however, fail to directly compare the three types of HMO physicians with the traditional fee-for-service physician, invoking instead the tacit assumption that physicians working in IPA model HMOs are sufficient proxies for non-HMO fee-for-service physicians. Thus, the strength of Mechanic's study is that it directly compares fee-for-service physicians with HMO physicians, while the strength of the Wolinsky and Corry study is that it identifies the different effects produced by various HMO organizational structures. To be sure, the strength of one study is the weakness of the other, and vice versa. Accordingly, although these studies address the same general issue, they utilize methods and data that do not permit direct comparisons.

In the present study we incorporate the best of the Mechanic and the Wolinsky and Corry studies to assess the impact of the organizational structure of medical practice on the amount of time physicians spend with their patients. To do so we pool cross-sectional data from two national studies of physicians and their medical practices. The pooled data set contains information on the medical practices of 5,133 physicians working in the conventional fee-for-service system, and 1,642 physicians working in HMOs. To control for variations in case-mix, we limit the analysis to a subset of 2,521 primary care physicians, for whom all of the appropriate information had been obtained. We then use multiple regression techniques to determine the impact of organizational structure on the amount of time that physicians spend with their patients overall, in the office, and in the hospital.

Data and Methods

The data on fee-for-service physicians are taken from the 14th annual Periodic Survey of Physicians (PSP-14M) conducted in 1980 by the Center for Health

Services Research and Development of the AMA. The PSP-14M questionnaire was mailed to a 5 per cent, single-stage, proportionately stratified (on medical specialty categories) random sample of nonfederal, office-based physicians engaged in active medical practice in the United States. The sample was taken directly from the AMA Physician Masterfile's enumeration of every physician, regardless of membership, in the United States. By using a series of four follow-up mailings at approximately monthly intervals, a useable completion rate of 50 per cent was obtained providing data on approximately 5,133 physicians.

An analytic comparison of the medical specialty, census division, age, and sex of responding physicians with the physician population revealed no substantive differences.[11] Although the distributions on these known characteristics are demonstrably similar, the comparability of the distributions of the respondents and the non-respondents on the unknown characteristics (e.g., the amount of time spent with patients) remains unknown. However, Barton and her colleagues[12,13] recently have shown that when a reasonable response rate is obtained from a homogeneous population like ours, the estimation of statistical associations such as those reported herein are "unlikely to be affected by major bias due to non-response." Moreover, the most frequent bias identified in the nonresponse literature is that the respondents generally are better educated than the nonrespondents.[14-16] Obviously, this potential bias is not applicable to our study of physicians. Finally, in a more clinical vein, Criqui and his colleagues have shown that when respondents and nonrespondents with similar distributions on exposure variables (e.g., demographic and occupational factors) are examined, comparable prevalence and relative risk estimates are obtained.[17,18] Accordingly, we assume that given our purpose, the high response rate, the homogeneity of the population, and the marked similarity between the known characteristics of the respondents and the nonrespondents, these data are not likely to be unduly distorted by nonresponse biases. Thus, the PSP-14M data provide an appropriate data source with which to empirically assess the impact of organizational structure on the amount of time that physicians spend with their patients in the traditional fee-for-service setting.

Comparable data on physicians practicing in HMOs was obtained in two stages. First, a letter was sent to all HMOs listed in the 1978 census of HMOs compiled by the federal Office of Health Maintenance Organizations (OHMOs).[19] The executive directors of each of the 203 HMOs enumerated in the census were asked to provide a list of all physicians participating in their plans. Ninety-seven of the HMOs responded and an additional 17 lists were obtained from the Register of Group Practice concurrently being assembled by the AMA. An analytic comparison of the response rates for the three major types of HMOs failed to indicate any nonresponse biases. Second, a 20 per cent sample then was selected by systematically choosing one out of every five physicians from each HMO physician list using a new random seed within each HMO. The HMO physician sample then was sent the standard PSP-14M questionnaire during 1980. Using four follow-up mailings to subsequently nonresponding physicians at approximately monthly intervals, a useable response rate of 50 per cent was obtained, resulting in a sample of 1,642 HMO physicians. Based on the most recent categorizations published by the OHMO in the *Federal Register*, responding physicians were classified as participating in group model (N = 294), staff model (N = 163), or IPA model (N = 1,160) HMOs. On all known criteria, which are admittedly few and not specific, the resulting sample of HMO physicians (PSP-14H) appears to be representative of the population of HMO physicians.[10] Thus, PSP-14H represents an appropriate data source with which to assess the impact of organizational structure on the amount of time that physicians in HMOs spend with their patients.

When taken together, the PSP-14M and PSP-14H physician surveys comprise the data set to be used in this paper (hereafter referred to as PSP-14). The pooled PSP-14 data set allows us to compare the impact of organizational structure on the amount of time that physicians spend with their patients in the fee-for-service setting directly with the amount of time that physicians in IPA model, group model, and staff model HMOs spend with their patients. Thus, these data permit us to make the direct comparisons between fee-for-service system physicians and HMO physicians that Mechanic made, as well as distinguishing between the different types of HMO settings examined by Wolinsky and Corry.

In this study the dependent variables are the numbers of minutes spent per patient visit overall, in the office, and in the hospital. These variables were calculated from information contained on the PSP-14 questionnaires concerning the total number of hours worked, the total number of patients seen, the total number of hours worked in the office, the total number of patients seen in the office, the total number of hours worked in the hospital, and the total number of patients seen in the hospital. In all cases, physicians were asked to indicate the total number of hours (or visits) in their most recent, complete week of practice.

Unfortunately, such retrospective self-reporting has a serious limitation and we would be remiss not to discuss it in some detail. This limitation recently has been addressed by Nelson et al.[20] with special reference to physicians. Using techniques logically consistent with classic multitrait-multimethod and triangulation approaches,[21–25] Nelson et al. compared the retrospective self-reports of 13 primary care physicians with checklists obtained from independent shadow-observers over a 1-week period. They failed to find close agreement between the two data sources concerning the frequency and duration of various tasks (e.g., the number and length of patient visits); physicians reported doing more tasks and doing them longer than the observers' records would indicate. As a result, the point estimates (i.e., the means) obtained by using the physicians' retrospective self-reports probably were overstated. This overstatement may have resulted from the physicians including in their actual patient contact estimates the amount of time that they spent reviewing charts, writing case histories, and attending grand rounds. In any event, Nelson et al. argue that the retrospective self-reporting method, at least among physicians, may not be valid for generating point estimates.

While the validity of the point estimates themselves may be questioned (e.g., the amount of time spent with patients may be overestimated), there is no empirical evidence to suggest that this tendency to overestimate is more or less pronounced across the different organizational settings in which physicians may practice. That is, physicians in staff model HMOs are no more or no less likely to over- or underestimate the amount of time that they spend with their patients than are solo practitioners or physicians in fee-for-service group practices. Thus, the differences between the point estimates for various groupings of physicians based on the organizational structure of their medical practices should be reliable and meaningful. Therefore, our use of these data for focusing on the differences in the amounts of time that physicians spend with their patients is indeed appropriate. Accordingly, we focus and report solely on the differences between the point estimates for physicians in the various organizational structures and not on the point estimates themselves.

The principal independent variables were dummies reflecting the organizational setting in where the physician practiced medicine, including: solo fee-for-service, group model HMOs, staff model HMOs, and IPA model HMOs. The dummy variable for the fee-for-service group practice category was omitted to preserve the singularity of the correlation matrix. Thus, the regression coefficient for each of the dummy variables indicates the difference between being in that particular organizational structure as opposed to being in fee-for-service group practice.

In addition to the principal independent variables measuring organizational structure, several others were included to control for a variety of effects. For example, in order to control for case-mix variations (even though we restricted the analysis to primary care physicians), we include dummy variables for internal medicine, pediatrics, and obstetrics-gynecology. The omitted category is that of general and family practice. Thus, the regression coefficients for each of these dummy variables indicates the difference between that particular medical specialty and general or family practice. Similarly, to control for variations in reimbursement methods we constructed a dummy variable that indicated whether or not the physician was paid for the majority of his or her practice on a straight salary basis.§ This partitions the effect of an individual physician's reimbursement from the more macro-level structural effects of the organizational reimbursement setting of his or her medical practice.

Finally, we also controlled for several other factors, including exposure to the HMO (not all HMO physicians work full time for the HMO), the physician's age (as a proxy for life-cycle stage), and the interaction of individual salary reimbursement and working in a staff model HMO (because the combination of a salaried organizational setting and being salaried oneself would be especially reinforcing). None of these exposure, age, and salary-staff model interaction terms, however, had any significant impact on the amount of time that physicians spend with their patients. We mention having controlled for these factors here only for the benefit of the reader.

Findings

Table 1 contains the regression coefficients, intercepts, and squared multiple correlation coefficients (r^2) obtained from the models predicting the amount of time physicians spend with their patients overall, in the office, and in the hospital. Coefficients significantly different from zero at the 0.05 level or beyond are indicated with an asterisk. In general, the model explained significant proportions of the variance in the amount and time physicians spend with their patients, although the predictive power was greater for the overall and office-visit indicators. The particular effects of the specialty, solo fee-for-service, HMO, and reimbursement characteristics are described below.

Specialty Characteristics

As indicated earlier, the regression coefficients for each of the three specialty dummies indicate the difference between that specialty and the control category of general or family practitioners. The significance of the specialty dummies indicates the importance of controlling for case-mix variation when assessing the amount of time physicians spend with their patients. That is, the data indicate that internists spend nearly 11 minutes more, overall, and nearly 10 minutes more in the

---

§ Although it would seem that this reimbursement method variable might create a collinearity problem when used in conjunction with the staff model HMO dummy, this clearly was not the case. The zero-order correlation between these two variables was only 0.203. This low correlation probably reflects the growing tendency of partners to incorporate and then arrange to have themselves placed on a salary by the corporation, thus minimizing their income tax liabilities. Because we simultaneously control for the effects of fee-for-service group practice (our reference category) as well as for the effects of staff model HMOs, this anomaly should neither cloud nor constrain a straightforward interpretation of the salaried reimbursement variable. That is, the fee-for-service group practice reference category should measure the effects of being in fee-for-service group practice (including being salaried in a fee-for-service group practice), and the staff model HMO dummy should measure the effects of being in a salaried reimbursement environment. Thus, the micro-level reimbursement variable should be measuring solely the effects of the individual physician's being salaried.

TABLE 1. Regression Coefficients Obtained From Models Predicting the Amount of Time That 2,521 Primary Care Physicians Spend With Their Patients Overall, in the Office, and in the Hospital, Per Visit.*

Independent Variables	Dependent Variables†		
	Overall	In the Office	In the Hospital
Intercept	23.30	20.11	37.90
Internal medicine	10.85*	9.68*	−3.65
Pediatrics	−0.80	−0.88	5.78
Obstetrics-gynecology	6.07*	0.95	18.70*
Solo F-F-S	3.57*	1.99*	11.47*
IPA model HMO	−0.49	−0.29	−1.44
Group model HMO	−3.19*	−3.58*	2.25
Staff model HMO	1.82	−0.81	14.67*
Salaried reimbursement	3.94*	2.99*	6.28
R^2	0.11*	0.13*	0.04*

* Coefficients significant at the 0.05 level or beyond.

† Measured in minutes per visit. Alternative transformations (including normal logarithms and polynomial expansions) failed to alter the findings, suggesting that minutes per visit is the appropriate measurement unit.

office with each of their patients, than general or family practitioners. That the coefficient for internal medicine on the hospital visits measure is not significant and has a negative sign probably reflects the fact that internists provide more extensive diagnostic services in their offices compared to more brief consultative services in hospitals. Pediatricians, on the other hand, do not spend significantly more or less time than general or family practitioners with their patients in any of the three settings. Finally, obstetricians-gynecologists spend 6 minutes more, overall, and nearly 19 minutes more in the hospital, than their general or family practice counterparts. This probably reflects the greater hospital or surgical orientation of obstetricians-gynecologists when compared to general or family practitioners.

Solo Fee-for-Service

Perhaps one of the more interesting findings portrayed in Table 1 is the significant and positive coefficients that the solo fee-for-service dummy produces on all three dependent variables, confirming Mechanic's earlier work.[8] When compared to their group fee-for-service counterparts, solo fee-for-service physicians spend nearly 4 minutes more with their patients overall, nearly 2 minutes more with their patients in the office, and over 11 minutes more with their patients in the hospital. According to Freidson, the consistently greater amounts of time spent by solo fee-for-service physicians with their patients may well reflect their greater sense of client dependency, as opposed to the greater colleague dependency felt by their group fee-for-service counterparts.[26] That is, Freidson argues that solo practitioners are more directly dependent on their patients for their patient flow (especially when in general or family practice) than their colleagues who band together in some form of collaborative practice arrangements. In essence, physicians in collaborative practice arrangements may receive patients through the secondary benefit of the brand-name effect of the group or on direct referral from their partners; physicians in solo practice do not enjoy the advantages of such secondary benefits or direct referral networks.

The HMO Characteristics

Three rather interesting patterns are reflected in the regression coefficients produced by the HMO organizational types.

First, the regression coefficients indicate that IPA model HMO physicians do not spend significantly more or less time with their patients in any setting than physicians in fee-for-service group practices. This is to be expected, since the IPA model HMO essentially is a sideline for its physicians and usually amounts to 10 per cent or less of their total medical practices. Therefore, although legally associated with an HMO, IPA physicians are not likely to be affected by its incentive structures. Second, the coefficients for group model HMO physicians provide some support for Mechanic's interpretation regarding the total and office visits measures. Physicians in group model HMOs spend significantly less time (about 3–4 minutes) than their counterparts in fee-for-service group practices. Thus, as Mechanic suggested, group model HMO physicians appear to adjust to increased patient demand by shortening the amount of time they spend with each patient, rather than increasing their work week as do their fee-for-service group practice counterparts. Finally, however, this trend is contradicted by the significantly greater amount of time (approximately 15 minutes) that staff model HMO physicians spend with their patients in the hospital. These data indicate that physicians in staff model HMOs, who generally are salaried employees, spend more time with their patients (at least in the hospital) than their fee-for-service group practice counterparts, who have a proprietary interest in generating practice incomes.

Reimbursement Characteristics

The regression coefficients for the salaried reimbursement method indicate that salaried physicians spend nearly 5 minutes more with their patients overall, and nearly 3 minutes more with their patients in the office than their nonsalaried counterparts. Like the effects of being in a staff model HMO, the effects of being reimbursed on a salaried basis support Wolinsky and Corry's argument and contradict Mechanic's hypothesis. Mechanic posited that salaried physicians would adjust to increased patient demand by spending less time with their patients, which clearly is not the case. Wolinsky and Corry, however, argued that in the absence of fiscal incentives encouraging physicians to see more different patients more often, salaried physicians would spend more time with their patients.

Summary

Unfortunately the findings reported above are not sufficient to resolve the controversy surrounding the amount of time that physicians in different organizational settings spend with their patients. While some of the findings support Mechanic's interpretation (i.e., that group model HMO physicians generally spend less time with their patients than physicians in fee-for-service group practices), others support Wolinsky and Corry's interpretation (i.e., salaried physicians spend more time with their patients than nonsalaried physicians). These data do, however, provide several valuable insights concerning the relationship between organizational structure and individual physicians' practices. First, the difference in the coefficients for the staff model HMO dummy and the salaried reimbursement variable indicate that there is, indeed, a difference between being salaried and working in a predominantly salaried environment. Moreover, these data indicate that the micro-level incentive (i.e., that the physician is on a salary) is more effective in altering individual physicians' medical practices than the macro-level incentive (i.e., merely working in an environment where most of one's colleagues are salaried). Second, it is important to control for the effects of case-mix when examining the impact of organizational structure on individual physicians' practices, as indicated by the substantial regression coefficients for the medical specialty dummies. Finally, the medical practices of physicians working in IPA model HMOs are not significantly different from those of their counterparts in fee-for-service group practices. Therefore, the use of data on physicians in IPAs as a proxy for data on fee-for-service physicians when making comparisons to data on physicians in staff and group model HMOs may well be appropriate.

Implications

Two implications for future research and policy may be drawn from the findings reported above. First, future research must explicitly identify and measure the specific incentives that, when bundled together, constitute the various organizational structures of medical practice. Until these specific incentives have been identified and measured, it will not be possible to determine exactly why medical practice varies from one organizational structure to the next.[6] Second, future research also must be based upon the analysis of longitudinal rather than cross-sectional data in order to determine whether there is a selection process that brings fundamentally different types of physicians to practice in the different organizational structures.[26] Although the cross-sectional data used herein, as well as that used by Mechanic, and by Wolinsky and Corry, have been useful in establishing the relationship between variations in organizational structures and differences in individual physicians' medical practices, clear causal inferences only can be made from longitudinal data.

References

1. Enthoven A. Consumer choice health plan. N Engl J Med 1978;298:650.

2. PL 93-641. The national health planning and resources development act of 1974. Washington, D.C.: U.S. Government Printing Office. 1974.

3. Blumberg M. Health status and health care use by type of private health coverage. Milbank Mem Fund Quarterly 1980;58(4):633.

4. Iglehart J. The federal government as venture capitalist: how does it fare? Milbank Mem Fund Q 1980;58(4):656.

5. Luft H. Assessing the evidence on HMO performance. Milbank Mem Fund Q 1980;58(4):501.

6. Wolinsky F. The performance of HMOs: an analytic review. Milbank Mem Fund Q 1980;58(4):537.

7. Mechanic D. General medical practice: some comparisons between the work of primary care physicians in the United States and England and Wales. Med Care 1972;10:402.

8. Mechanic D. The organization of medical practice and practice orientations among physicians in prepaid and nonprepaid primary care settings. Med Care 1975;13:189.

9. Held P, Reinhardt U. Analysis of economic performance in medical group practices. Princeton: Mathematica Policy Research, 1979.

10. Wolinsky F, Corry B. Organizational structure and medical practice in health maintenance organizations. In: Goldfarb D, ed. Profile of medical practice, 1981. Chicago: American Medical Association, 1981.

11. Goldfarb D, ed. Profile of medical practice, 1981. Chicago: American Medical Association, 1981.

12. Barton J, et al. Characteristics of respondents and non-respondents to a mailed questionnaire. Am J Public Health 1980;70(8):823.

13. Belanger C, et al. The nurses' health study. Am J Nurs 1978;78(9):1038.

14. Suchman E, McCandless B. Who answers questionnaires? J Appl Psychol 1940;24(7):758.

15. Edgerton H, et al. Objective differences among various types of respondents to a mailed questionnaire. Am Sociol Rev 1947;12(4):435.

16. Reuss C. Differences between persons responding and not responding to a mailed questionnaire. Am Sociol Rev 1943;8(4):433.

17. Criqui M, et al. Differences between respondents and non-respondents in a population-based cardiovascular disease study. Am J Epidemiol 1978;108(4):367.

18. Criqui M. Response bias and risk ratios in epidemiologic studies. Am J Epidemiol 1979;109(4):394.

19. U.S. Department of Health, Education, and Welfare. Census of prepaid plans, 1978. Washington, D.C.: U.S. Government Printing Office, 1978.

20. Nelson E, et al. A study of the validity of the task inventory method of job analysis. Med Care 1975;13(2):104.

21. Campbell D, Fiske D. Convergent and discriminant validation by the multitrait, multimethod matrix. Psychol Bull 1969;56(1):81.

22. Denzin N. The research act. 2nd ed. New York: McGraw-Hill, 1978.

23. Campbell D. Social attitudes and other acquired behavioral dispositions. In: Koch S, ed. Psychology: a study of science, vol 6. New York: McGraw-Hill, 1963.

24. Webb E. Unconventionality, triangulation and inference. In: Proceedings of the 1966 invitational conference on testing problems. Princeton: Educational Testing Service, 1966.

25. Webb E, et al. Unobtrusive measures: nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

26. Freidson E. The profession of medicine. New York: Dodd, Mead, 1970.